

## **Orbiting Carbon Observatory-2 & 3 (OCO-2 & OCO-3)**



### **Level 2 Full Physics Retrieval Algorithm Theoretical Basis**

Version 2.0 Rev 3  
December 1, 2020

National Aeronautics and  
Space Administration



Jet Propulsion Laboratory  
California Institute of Technology  
Pasadena, California

# **ORBITING CARBON OBSERVATORY**

## **(OCO) - 2**

### **LEVEL 2 FULL PHYSICS ALGORITHM**

#### **Theoretical Basis Document**

**Version 3.0 – Rev 1**  
**January 15, 2021**

David Crisp	Jet Propulsion Laboratory
Chris O'Dell	Colorado State University
Annmarie Eldering	Jet Propulsion Laboratory
Brendan Fisher	Jet Propulsion Laboratory
Fabiano Oyafuso	Jet Propulsion Laboratory
Vivienne Payne	Jet Propulsion Laboratory
Brian Drouin	Jet Propulsion Laboratory
Geoff Toon	Jet Propulsion Laboratory
Joshua Laughner	California Institute of Technology
Peter Somkuti	Colorado State University
Greg McGarragh	Colorado State University
Aronne Merrelli	University of Wisconsin – Madison
Robert Nelson	Jet Propulsion Laboratory
Michael Gunson	Jet Propulsion Laboratory
Christian Frankenberg	California Institute of Technology
Gregory Osterman	Jet Propulsion Laboratory
Hartmut Boesch	University of Leicester
Linda Brown	Jet Propulsion Laboratory
Rebecca Castano	Jet Propulsion Laboratory
Mathew Christi	Colorado State University
Brian Connor	BC Consulting Limited
James McDuffie	Jet Propulsion Laboratory
Charles Miller	Jet Propulsion Laboratory
Vijay Natraj	Jet Propulsion Laboratory
Denis O'Brien	Colorado State University
Igor Polonski	Colorado State University
Mike Smyth	Jet Propulsion Laboratory
David Thompson	Jet Propulsion Laboratory
Robert Granat	Jet Propulsion Laboratory

**Document History:**

Version	Revision	Date	Description/Comments
1.0	1		Initial version of OCO ATBD
1.0	2	4/18/2009	Final version of OCO ATBD
1.0	3	10/15/2010	First version of OCO-2 ATBD
1.0	4	11/10/2010	Final pre-external review version, minor fixes
2.0	0	4/17/2014	Prelaunch version for OCO-2
2.0	1	2/21/2015	Draft postlaunch version for OCO-2
2.0	2	3/30/2015	Accompanying V6 and V6R public data release
2.0	3	1/2/2019	Edited to include OCO-3
3.0	1	1/15/2021	Updated for OCO-2 v10 and OCO-3 vEarly data

The research described in this document was carried out at the Jet Propulsion Laboratory,  
California Institute of Technology, under a contract with the  
National Aeronautics and Space Administration.  
Copyright ©2021. All rights reserved.

**TABLE OF CONTENTS**

<b>1</b>	<b>Introduction .....</b>	<b>1</b>
1.1	Purpose .....	1
1.2	Scope .....	1
<b>2</b>	<b>Experiment Overview .....</b>	<b>2</b>
2.1	Objectives .....	2
2.2	Measurement Approach.....	3
2.2.1	Observation strategies common to OCO-2 and OCO-3.....	4
2.2.2	OCO-2 glint/nadir/target observation strategy .....	4
2.2.3	OCO-3 Glint/Nadir/Target/SAM Strategy .....	7
2.3	Data Product Delivery .....	9
<b>3</b>	<b>Algorithm Description .....</b>	<b>10</b>
3.1	Algorithm Overview .....	10
3.2	Forward Model .....	15
3.2.1	Optical Properties .....	17
3.2.2	Solar Spectrum Model.....	31
3.2.3	Radiative Transfer .....	33
3.2.4	Instrument Model .....	38
3.3	State Vector .....	40
3.3.1	Basic Description of State Vector .....	40
3.3.2	The <i>a Priori</i> State Vector and Covariance Matrix .....	41
3.4	Jacobians.....	52
3.5	Inverse Method .....	53
3.5.1	Formulation and Implementation .....	53
3.5.2	Goodness of Spectral Fit .....	55
3.6	X <sub>CO2</sub> , Characterization, and Error Analysis.....	56
3.6.1	Pressure Weighting Function .....	56
3.6.2	X <sub>CO2</sub> .....	57
3.6.3	The Column Averaging Kernel $a_{CO2}$ .....	57
3.6.4	Smoothing and Interference due to the State Vector $\tilde{a}_c$ .....	58
3.6.5	Correlation of X <sub>CO2</sub> with non-CO <sub>2</sub> State Vector Elements $\tilde{s}_1$ .....	59
3.6.6	Components of X <sub>CO2</sub> Variance .....	59
<b>4</b>	<b>Algorithm Specification .....</b>	<b>61</b>
4.1	Data System Context .....	61
4.2	Level 2 Preprocessing.....	61
4.2.1	ABO2 Preprocessor.....	61
4.2.2	IMAP Preprocessor (and SIF Retrievals).....	62
4.3	Level 2 Full Physics Inputs .....	62
4.3.1	Requirements on Inputs.....	63
4.3.2	Attributes of Input Products .....	63
4.4	Level 2 Full Physics Outputs.....	64

<b>4.5</b>	Description of Full Physics Major Code Sections .....	65
<b>4.6</b>	Lua Configuration.....	66
<b>4.7</b>	Level 2 Processing PGEs.....	67
<b>4.7.1</b>	GEOS FP-IT Resampler.....	67
<b>4.7.2</b>	L2 A-Band Preprocessing PGE.....	68
<b>4.7.3</b>	L2 IMAF-DOAS Preprocessing PGE .....	68
<b>4.7.4</b>	Sounding Selection PGE .....	68
<b>4.7.5</b>	L2 FP Algorithm .....	68
<b>4.7.6</b>	FP Diagnostic Product Generator.....	69
<b>4.7.7</b>	FP Standard Product Generator.....	69
<b>4.7.8</b>	Level 2 PGE Multiplicity Summary.....	69
<b>4.8</b>	Error Handling.....	70
<b>4.9</b>	Commercial off-the-Shelf Components.....	70
<b>4.10</b>	Quality Assessment and Recording .....	70
<b>5</b>	References .....	71
<b>6</b>	Acronyms .....	81

## LIST OF FIGURES

Figure 2-1. Nadir, glint, and target observations. (a) Nadir observations are acquired over the sunlit hemisphere at latitudes where the surface solar zenith angle is $<85^\circ$ . On all orbits except downlink orbits, as the Observatory passes over the northern terminator, it pitches up to point the instrument aperture at the Sun for solar radiometric calibrations. (b) Glint observations are made at latitudes on the sunlit hemisphere where the solar zenith angle of the glint spot is less than $75^\circ$ . (c) For target observations, the spacecraft points the instrument at a stationary surface target as it flies over. A small-amplitude sinusoidal oscillation in the pitch axis is superimposed on the nominal pointing to scan the spectrometer slit across the target. ....	5
Figure 2-2. Examples of footprint positions for a selection of 6 positions through an OCO-2 orbit across a span in latitude from 70S to 70N. All plots show the relative footprint positions, by converting the geolocated latitude and longitude to approximate linear distance in kilometers, and subtracting the boresight center as approximated by the mean of all 8 footprint positions. An arbitrary offset of 8 km in the N-S direction is applied to each frame. Footprints 1 and 8 are marked. (a) on left, shows a nadir (ND) and glint (GL) orbit from the early mission before the spacecraft yaw offset was commanded. (b) on right, shows a nadir and glint orbit later in the mission after the spacecraft yaw offset was commanded. ....	6
Figure 2-3. OCO-2 O <sub>2</sub> A-Band radiances for an OCO-2 SAM taken over Buenos Aires, Argentina at 17:22 UTC on 22 September 2019, showing the spatial coverage possible with a typical SAM data collection. ....	8
Figure 2-4. Flow diagram of data processing strategy from the raw instrument data through L2 and lite files. The data products are shown in grey, while the product generation executives (PGEs) are shown in blue. ....	9
Figure 3-1. Viewing geometry and a few possible optical paths traversed by solar photons that are recorded by the instrument. ....	11
Figure 3-2. Spectral dependence of the absorption coefficients of CO <sub>2</sub> at wavelengths within the strong CO <sub>2</sub> band near 2.08 $\mu\text{m}$ are shown for pressures near 0.1, 100, and 600 hPa (mbar). ....	13
Figure 3-3. Flow diagram for the L2 retrieval algorithm. ....	15
Figure 3-4. Example mean spectral residual for high airmass ( $\text{sza} > 75^\circ$ ) for the O <sub>2</sub> A band using the same set of TCCON spectra used in the analysis of the absco 5.1 update for CO <sub>2</sub> (Oyafuso [2017]). Spectroscopic residuals are evident near the band head, between the P and R branch and in between doublets in the P branch. ....	20
Figure 3-5. Residuals for high airmass ( $\text{sza} > 75^\circ$ ) atmospheric spectra at 1.6 $\mu\text{m}$ . Data set consists of 164 spectra and several problematic water lines have been masked out as described in Oyafuso et al. [2017]. ....	21
Figure 3-6. Residuals for high airmass atmospheric spectra ( $\text{sza} > 75^\circ$ ) at 2.06 $\mu\text{m}$ . See text associated with Figure 3-5. Figure adapted from Oyafuso et al. [2017]. ....	22
Figure 3-7. Optical depth contributions of different molecules in the O <sub>2</sub> A band for a US standard atmosphere. ....	23
Figure 3-8. Optical depth contributions for different gases in the 1.6 $\mu\text{m}$ CO <sub>2</sub> band for a tropical standard atmosphere. ....	23
Figure 3-9. Optical depth contributions from different molecules in the 2.06 $\mu\text{m}$ CO <sub>2</sub> band for a tropical standard atmosphere. ....	24

Figure 3-10. Independent phase matrix elements for the cloud ice particles used in the L2 retrieval algorithm. Shown are both the exact phase functions, as well as the M=5000 delta-M. 28

Figure 3-11. Aerosol and cloud optical properties assumed in the L2 code. Left: Extinction Efficiency relative to that at 755 nm. Middle: Single Scattering Albedo. Right: Asymmetry Parameter. DU: Dust, SS: Sea Salt, BC: Black Carbon, OC: Organic Carbon, SO: Sulfate, WC: Water Cloud, IC: Ice Cloud. The three OCO-2 bands are demarcated by the dashed vertical lines. .... 32

Figure 3-12. Percent errors in OCO-simulated TOA radiances due to the LSI method for four different cases. Each case includes one or multiple clouds. The three OCO bands are shown in blue (O<sub>2</sub> A), green (weak CO<sub>2</sub>), and red (strong CO<sub>2</sub>). Taken from O'Dell [2010], Figure 8. .... 36

Figure 3-13. Illustration of non-uniform sampling approach. Top: Training. From a full monochromatic spectrum (black diamonds) certain spectral points are selected for RT calculations (blue squares). The remaining points are calculated from linear interpolation of the selected points. Bottom: Interpolation errors in the non-selected points for different interpolation schemes. Linear interpolation is used because its errors are comparable to higher-order interpolation schemes. .... 37

Figure 3-14. Geometry of the reference plane for polarization, i.e., the local meridian plane. This plane is formed by the local normal unit vector and the ray from the target FOV to the satellite. The direction of propagation of measured light is the latter. Two example polarizations are shown depicting light polarized both parallel (||) and perpendicular (⊥) to this plane. The polarization angle of the instrument is defined as the angle between the axis of accepted polarization and the reference plane for polarization. .... 39

Figure 3-15. *A priori* CO<sub>2</sub> correlation matrix (in colors), together with the 1-sigma errors for each pressure level. .... 42

Figure 3-16. Prior aerosol profiles uses the L2FP retrieval. .... 44

Figure 3-17. Global distribution of modeled aerosol type for July 2009. (Panel A) dominant type; (Panel B) second-most dominant type. .... 45

Figure 3-18. Global distribution of modeled aerosol optical depth for July 2009. (Panel A) dominant type; (Panel B) second-most dominant type; (Panel C) fractional contribution of the two dominant types to total AODs. .... 46

Figure 3-19. Example of first two leading eigenvectors for current GOSAT retrievals in the O<sub>2</sub> A band (left) and strong CO<sub>2</sub> band (right). .... 51

Figure 4-1. Data flow through the L2 FP software components. .... 66

Figure 4-2. L2 Processing overview. .... 67

**LIST OF TABLES**

<b>Table 3-1.</b> Inputs and outputs of the forward model. ....	16
<b>Table 3-2.</b> ABSCO v. 5.1 parameter sources. ....	25
<b>Table 3-3.</b> Density weightings and RH values chosen for the L2 algorithm. ....	29
<b>Table 3-4.</b> Description of state vector elements. ....	41
<b>Table 3-5.</b> Aerosol prior profile quantities and <i>a priori</i> uncertainties. ....	45
<b>Table 3-6.</b> Surface state variables and <i>a priori</i> uncertainties. ....	48
<b>Table 3-7.</b> Description of <i>outcome</i> values. ....	53
<b>Table 3-8.</b> Inverse method products recorded with each sounding. ....	57
<b>Table 4-1.</b> L1B Science product attributes. ....	63
<b>Table 4-2.</b> L2Met product attributes. ....	63
<b>Table 4-3.</b> L2 Single Sounding product attributes. ....	64
<b>Table 4-4.</b> L2Dia product attributes. ....	64
<b>Table 4-5.</b> L2Std product attributes. ....	65
<b>Table 4-6.</b> Daily L2 PGE instances. ....	69



# **1 Introduction**

## **1.1 Purpose**

This algorithm theoretical basis document (ATBD) describes the algorithms used to retrieve the column-averaged CO<sub>2</sub> dry air mole fraction,  $X_{CO_2}$ , and other quantities included in the Level 2 (L2) Product from the spectra collected by the National Aeronautics and Space Administration (NASA) Orbiting Carbon Observatory-2 (OCO-2) and Orbiting Carbon Observatory-3 (OCO-3) missions. In particular, this document identifies sources of input data, which are required for retrievals; describes the physical theory and mathematical background underlying the use of this information in the retrievals; includes implementation details; and summarizes the assumptions and limitations of the adopted approach.

## **1.2 Scope**

This document starts with a brief overview of the OCO-2 and OCO-3 instruments, their measurement approaches and primary data products. It then describes the Level 2 retrieval algorithms used to estimate  $X_{CO_2}$ , solar induced chlorophyll fluorescence (SIF) and other geophysical quantities from these measurements, which constitute the version 10 Level 2 Full Physics (FP) Products that are routinely delivered to the Goddard Earth Sciences Data and Information Services Center (GES DISC) by the OCO-2/OCO-3 Science Data Operations System (SDOS). It then describes the pre-processors and screening algorithms used to initiate the surface/atmosphere state properties and to identify and screen soundings that are contaminated by optically-thick clouds and aerosols or other artifacts. The overall architecture of the data processing pipeline is outlined in Section 4. References for publications cited in the text are given in Section 5. Section 6 provides a list of acronyms and their definitions.

## 2 Experiment Overview

### 2.1 Objectives

The Orbiting Carbon Observatory was the first NASA mission designed to collect space-based measurements of atmospheric carbon dioxide ( $\text{CO}_2$ ) with the precision, accuracy, resolution, and coverage needed to characterize the processes controlling its buildup in the atmosphere [Crisp et al., 2004; 2007]. After a launch mishap, which prevented the original OCO mission from reaching orbit, the Orbiting Carbon Observatory-2 (OCO-2) mission was formulated to meet the original OCO science objectives. As part of the OCO-2 development effort, a flight spare spectrometer was built. Once OCO-2 was successfully launched, the spare spectrometer became the heart of the OCO-3 payload, which is now returning  $X_{\text{CO}_2}$  measurements from the International Space Station (ISS) [Eldering et al., 2018; Taylor et al., 2020].

Fossil fuel combustion, heavy industry, agriculture, forestry, and other land use change are now adding more than 42 billion tons of carbon dioxide ( $\text{GtCO}_2$ ) to the atmosphere each year [Friedlingstein et al. 2019]. Since the beginning of the industrial age, these and other human activities have increased the globally averaged atmospheric  $\text{CO}_2$  concentration by almost 50%, from less than 277 parts per million (ppm) in 1750 to about 415 ppm today. These anthropogenic  $\text{CO}_2$  emissions are superimposed on an active natural carbon cycle that regulates  $\text{CO}_2$  through photosynthesis and respiration [Beer et al. 2010] and its solubility in the ocean [Ciais et al. 2013]. Ground-based measurements from a global network of about 150 stations indicate that the natural carbon cycle emits almost 20 times as much  $\text{CO}_2$  into the atmosphere as human activities each year, but then reabsorbs a comparable amount, along with more than half of the anthropogenic emissions. They also show that the resulting net atmospheric  $\text{CO}_2$  growth rate is correlated with large-scale temperature and precipitation anomalies, like those associated with major El Niño events and large volcanic stratospheric aerosol injections. However, these ground-based measurements do not have the resolution or coverage needed to identify and quantify the sources emitting  $\text{CO}_2$  into the atmosphere or the natural sinks absorbing it at the surface, frustrating efforts to predict how the atmospheric  $\text{CO}_2$  growth rate might change as the carbon cycle responds to climate change.

Recent advances in remote sensing technologies are providing new tools for measuring  $\text{CO}_2$  with much greater resolution and coverage from space-based platforms. High-resolution spectra of reflected sunlight within molecular oxygen ( $\text{O}_2$ ) and  $\text{CO}_2$  bands can be analyzed to yield precise, spatially resolved estimates of the column-averaged  $\text{CO}_2$  dry air mole fraction,  $X_{\text{CO}_2}$ . The principal challenge of this approach is the need for unprecedented levels of precision and accuracy. Although natural processes and human emissions can change the atmospheric  $\text{CO}_2$  mixing ratio by as much as 8% near the surface (>30 ppm out of the ~400 ppm background), the amplitude of these variations decreases rapidly with altitude, such that  $X_{\text{CO}_2}$  variations rarely exceed 2% (8 ppm) on regional to global scales. East-west variations are typically no larger than 0.3 to 0.5% [1–2 ppm; Miller et al., 2007]. Because of this, modeling studies show that space-based measurements of  $X_{\text{CO}_2}$  can substantially improve our understanding of surface fluxes only if they have the accuracy, precision, coverage, spatial resolution, and temporal sampling needed to describe  $X_{\text{CO}_2}$  variations with amplitudes no larger than 0.3 to 0.5% (1 to 2 ppm) on scales ranging from <100 km over continents to ~1000 km over the ocean [Rayner and O’Brien, 2001; Miller et al., 2007; Baker et al., 2010; Chevallier et al., 2014; 2019].

To resolve  $\text{CO}_2$  fluxes on spatial scales ranging from <100 to ~1000 km, data must be collected at higher resolution to discriminate natural sinks from nearby sources. A small

sampling footprint also helps to ensure that some cloud-free soundings can be obtained even in partially cloudy regions, since the probability of measuring a cloud-free scene is inversely proportional to footprint size. A small sounding footprint is also needed to quantify CO<sub>2</sub> emissions from discrete point sources, such as individual power plants or cities, because the minimum detection limit (measured in kg of CO<sub>2</sub>) associated with a given concentration change (e.g., a 1-ppm variation in X<sub>CO2</sub>) is inversely proportional to the area of the footprint.

## 2.2 Measurement Approach

The OCO-2 and OCO-3 missions carry nearly identical, three-channel, high-resolution imaging grating spectrometers. These instruments are described in detail in the OCO-2 and OCO-3 L1B ATBD and in Crisp et al. [2017]. We have included only a brief summary here, focusing on those aspects which affect interpretation of the L2 retrieval products.

The OCO-2 and OCO-3 spectrometers make coincident, bore sighted measurements of reflected sunlight in the short-wave-infrared (SWIR) CO<sub>2</sub> bands centered near 1.61 and 2.06  $\mu\text{m}$  and in the near-infrared (NIR) molecular oxygen (O<sub>2</sub>) A band at 0.765  $\mu\text{m}$  [Crisp et al., 2017]. Simultaneous measurements from these three spectral regions are combined to define a single “sounding.” Each sounding is analyzed with remote-sensing retrieval algorithms to produce an estimate of X<sub>CO2</sub> for the atmospheric path between the Sun, the reflecting surface, and the instrument aperture. Measurements of absorption within the weak CO<sub>2</sub> band near 1.61  $\mu\text{m}$  and the stronger CO<sub>2</sub> band near 2.06  $\mu\text{m}$  provide information about the CO<sub>2</sub> column abundance. Measurements of absorption in the O<sub>2</sub> A-band provide constraints on the surface pressure and uncertainties in the atmospheric optical path length introduced by cloud and aerosol scattering and pointing errors. The OCO-2 and OCO-3 spectrometers collect spectra in eight contiguous spatial footprints across a narrow field of view (FOV) at 1/3 second intervals, yielding 24 soundings per second with footprints areas < 3 km<sup>2</sup> along a < 10 km wide ground track.

In addition to X<sub>CO2</sub>, the O<sub>2</sub> A-band channels of the OCO-2 and OCO-3 spectrometers return precise estimates of solar-induced chlorophyll fluorescence (SIF). This fluorescence must be quantified and corrected to ensure accurate estimates of X<sub>CO2</sub> [Frankenberg et al. 2012, 2011a]. SIF is also a functional proxy for terrestrial gross primary productivity (GPP; Frankenberg et al. 2011b; Guanter et al. 2012; Joiner et al. 2013; Köhler et al. 2015; Sun et al. 2017; 2018; Magney et al. 2019). SIF-based estimates of GPP are being combined with X<sub>CO2</sub> in atmospheric inverse models to provide new constraints on CO<sub>2</sub> uptake by the land biosphere and new insights into carbon-climate feedbacks [Liu et al. 2017; Palmer et al. 2019; Crowell et al., 2019].

The OCO-2 and OCO-3 spectrometers have approximately equal sized fields of view at the surface for each of the eight footprints. The cross-slit width is approximately 1.3 km at nadir, with 2.3 km in footprint length along-track for both instruments. The footprint shapes depend on slit orientation relative to the along-track direction, as described in the following sections describing the observation strategies.

The OCO-2 spacecraft flies in formation with the 705-km Afternoon Constellation (A-Train). This 705-km altitude, 98.8-minute, Sun-synchronous orbit has a 98.2-degree inclination, a 1:36:30 PM mean ascending equator crossing time, and a 16-day (233 orbit) ground track repeat cycle. OCO-2 follows a ground track that crosses the equator 217.3 km east of the World Reference System-2 ground track that is followed by the NASA Aqua platform. This orbit was originally chosen such that OCO-2 overflies the CloudSat radar and CALIPSO lidar ground footprints. Those two spacecraft left the A-Train to form the C-Train in September of 2018. OCO-2 remained in its A-Train orbit because its early afternoon mean local time is better suited

for acquiring observations of the absorption of reflected sunlight by CO<sub>2</sub> and O<sub>2</sub> because the Sun is high, maximizing the available signal. It also facilitates coordinated calibration and validation campaigns with other A-Train instruments, and synergistic use of OCO-2 data with that from other A-Train platforms. For routine science operations, OCO-2 points the spectrometer's FOV at the local nadir or near the "glint spot," where sunlight is specularly reflected from the surface. Nadir observations yield higher spatial resolution over land, while glint measurements have greater sensitivity over ocean. Each month, OCO-2 returns ~5.5 million soundings that are sufficiently cloud-free to yield full-column estimates of X<sub>CO2</sub> with single-sounding precisions near 0.5ppm and accuracies <1 ppm at solar zenith angles as large as 70° [Eldering et al. 2017].

The OCO-3 spectrometer was installed on the International Space Station (ISS) in May 2019 for a planned 3-year mission lifetime. While OCO-2 collects global X<sub>CO2</sub> and SIF measurements from a 1:30 p.m. sun-synchronous, near-polar orbit, OCO-3 collects these observations from dawn to dusk from the low-inclination (51°) ISS orbit. OCO-3 also includes a fast, two-axis pointing mirror assembly (PMA) enabling nadir, glint and target observations, like those from OCO-2. In addition, the OCO-3 PMA enables a new observing mode called Snapshot Area Maps (SAMs) that return X<sub>CO2</sub> and SIF measurements over targets as large as 80 km × 80 km. By combining OCO-2 and OCO-3 observations, we expect to gain a better understanding of both diurnal variations in CO<sub>2</sub> fluxes and urban emissions.

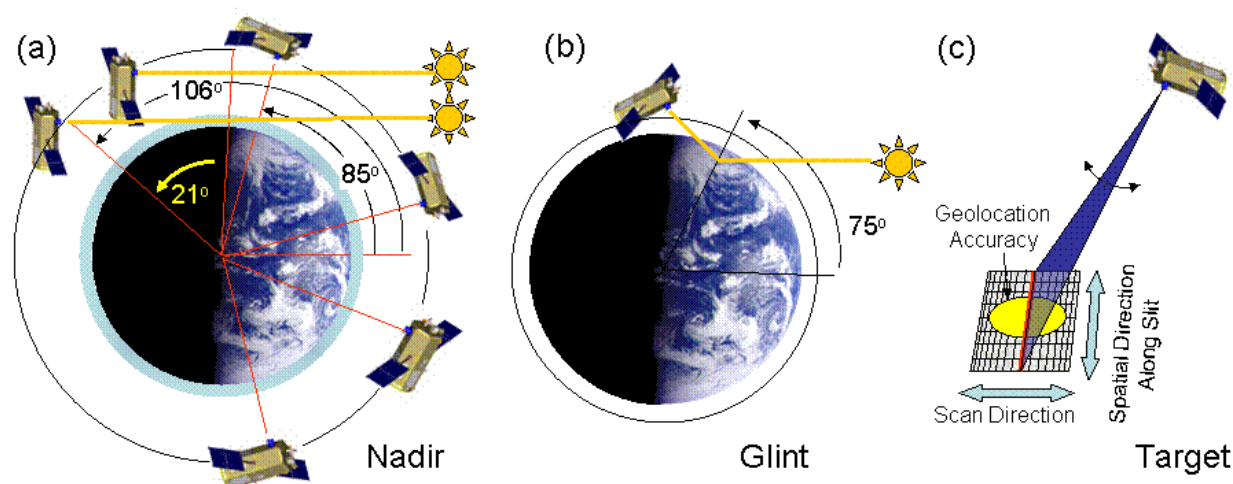
### **2.2.1 Observation strategies common to OCO-2 and OCO-3**

Both OCO-2 and OCO-3 collect science observations while viewing the local nadir and the apparent glint spot at the surface (Figure 2-1) [Crisp et al., 2007; Crisp, 2008]. Nadir observations are collected along the ground track just below the spacecraft. These measurements provide the highest spatial resolution, and yield more useable soundings in in partly cloudy regions and in regions with highly variable topography. However, nadir observations have limited SNR over dark ocean or ice-covered surfaces at high latitudes. Glint observations are collected while viewing a point between the local nadir the bright "glint" spot, where sunlight is specularly reflected from the Earth's surface. Glint observations provide much greater SNR or dark ocean surfaces.

The same rate of data sampling is used for all observations. The instrument collects eight adjacent, spatially resolved samples every 0.333 seconds (24 samples per second). At this data collection rate, the instruments collect ~400 soundings per degree of latitude as it travels from pole to pole, or about a million soundings over the sunlit hemisphere every day. Clouds, aerosols, and other factors will reduce the number of soundings available for X<sub>CO2</sub> retrievals, but the small sounding footprint ensures that some data will be sufficiently cloud free on regional scales at monthly intervals.

### **2.2.2 OCO-2 glint/nadir/target observation strategy**

For each sunlit portion of the orbit, OCO-2 will collect either glint or nadir geometry. The observatory does not switch between these two modes during one orbit. Target observations can be collected during either glint or nadir orbits, by switching modes briefly to collect the target data. OCO-2 Target observations are collected while pointing the instrument boresight at a stationary surface target as the satellite flies overhead. To ensure that the target is not missed, the slit is oriented perpendicular to the ground track and a small-amplitude ( $\pm 0.23^\circ$ ) sinusoidal oscillation along the ground track is superimposed on the boresight to scan the slits back and forth across the nominal target location as the instrument collects data (Figure 2-1c). This approach facilitates the collection of hundreds to thousands of soundings in a rectangular area



**Figure 2-1.** Nadir, glint, and target observations. (a) Nadir observations are acquired over the sunlit hemisphere at latitudes where the surface solar zenith angle is  $<85^\circ$ . On all orbits except downlink orbits, as the Observatory passes over the northern terminator, it pitches up to point the instrument aperture at the Sun for solar radiometric calibrations. (b) Glint observations are made at latitudes on the sunlit hemisphere where the solar zenith angle of the glint spot is less than  $75^\circ$ . (c) For target observations, the spacecraft points the instrument at a stationary surface target as it flies over. A small-amplitude sinusoidal oscillation in the pitch axis is superimposed on the nominal pointing to scan the spectrometer slit across the target.

surrounding a stationary target as the spacecraft flies overhead. Target observations are used to collect data over surface calibration and validation sites, and to provide greater spatial coverage and resolution of compact targets, such as large urban areas.

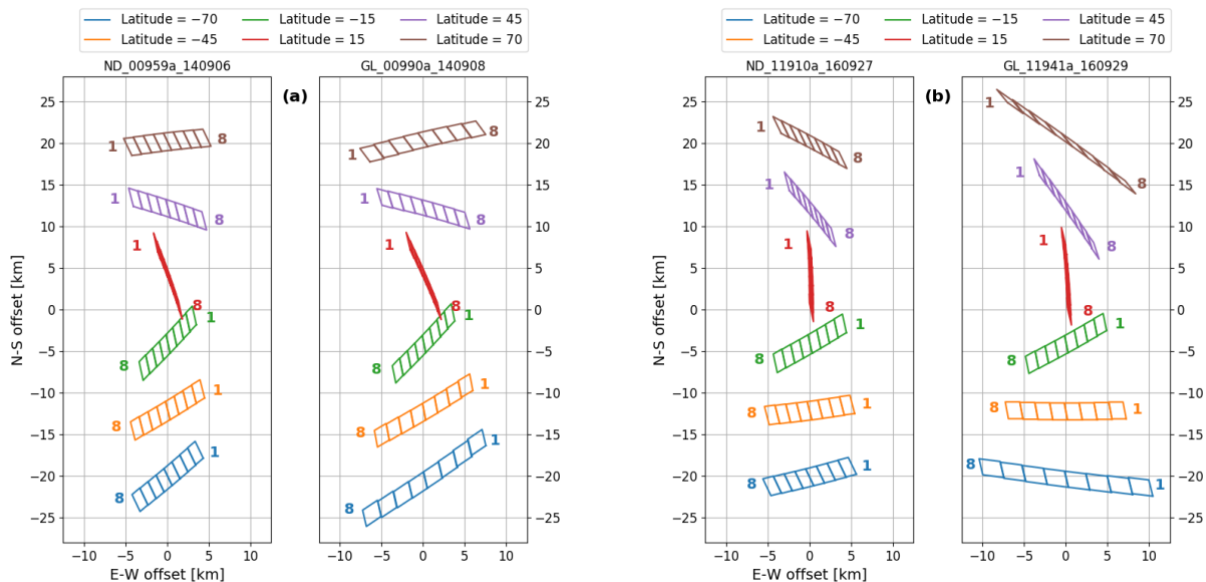
For OCO-2, nadir observations are collected at all locations where the surface solar zenith angle is less than  $85^\circ$ . Glint soundings are collected at all latitudes where the surface solar zenith angle is less than  $75^\circ$ . Target observations are collected over surface targets that are within  $61^\circ$  of the local spacecraft nadir along the orbit track and spacecraft viewing angles between  $30^\circ$  west of the ground track and  $5^\circ$  east of the ground track. When the target is near the ground track, a single pass can last up to nine minutes, providing 12000 to 24000 soundings in the vicinity of the target. This large number of soundings reduces the impact of random errors and provides opportunities to identify spatial variability in the  $X_{CO_2}$  field near the target.

While sunlight incident at the top of Earth's atmosphere is not polarized, scattering by the atmosphere and surface can introduce varying amounts of polarization in the observed reflected sunlight. The degree of polarization can vary substantially but is typically greatest at the Brewster angle over the ocean ( $\sim 53^\circ$  solar incidence angle) or near the terminators, where molecular (Rayleigh) scattering by the atmosphere introduces significant wavelength-dependent polarization. While the degree of polarization is typically unknown, the angle of polarization can be derived from the illumination and observing geometries. For example, if we define a "principal plane" that includes the sun, the surface footprint, and the instrument aperture, the light polarized parallel to this plane (e.g., the P polarization) is much more strongly attenuated than the light polarized perpendicular to this plane (e.g., the S polarization; Crisp et al. [2008]).

Since OCO-2 has a linear polarizer in the instrument optical path, in situations where the reflected sunlight is strongly polarized, the measured signal level will be strongly modulated by the relative orientation of the instrument to the principal plane. For OCO-2, the observing geometry was then chosen to impose a fixed polarization angle between the slit orientation and

the principal plane for glint and nadir observations. In order to keep this orientation fixed through an orbit, while the sun – footprint – sensor geometry is changing, the slit must rotate relative to the instrument ground track. The result is that the 8 footprints in a single frame will appear to rotate during an orbit. Figure 2-2a shows the apparent movement of the footprints around the boresight center at specific footprint latitudes that span a typical orbit. Note that the footprints in a glint orbit will be larger at higher latitudes, due to the larger observation zenith angles at these positions. Before 28 October 2014 (orbit 1722), all glint and nadir observations were collected with the long axis of the spectrometer slits oriented perpendicular to the principal plane. This observing geometry yielded adequate SNR over land and over ocean at low solar zenith angles, but provided too little signal for ocean glint observations collected near the Brewster angle ( $\sim 53^\circ$ ), where the component of the reflected sunlight polarized in the direction parallel to the principal plane (P polarization) is strongly attenuated. On 28 October 2014, the glint observing geometry was modified to acquire observations with the long axis of the spectrometer slits rotated  $30^\circ$  (clockwise from above) about the center of the instrument boresight. With this  $30^\circ$  rotation, the continuum SNR over the ocean increased to values between 200 and 600 at latitudes spanning the Brewster angle.

For nadir observations, OCO-2 initially continued to collect data with the long axis of the slits oriented perpendicular to the principal plane because these observations returned adequate signal only over land, where the reflected sunlight is not strongly polarized. However, it was soon discovered that transitions between nadir and glint observations changed the solar illumination geometry of the OCO-2 spacecraft bus enough to introduce variations in the instrument thermal environment that contributed calibration errors in the data. On 12 November



**Figure 2-2.** Examples of footprint positions for a selection of 6 positions through an OCO-2 orbit across a span in latitude from 70S to 70N. All plots show the relative footprint positions, by converting the geolocated latitude and longitude to approximate linear distance in kilometers, and subtracting the boresight center as approximated by the mean of all 8 footprint positions. An arbitrary offset of 8 km in the N-S direction is applied to each frame. Footprints 1 and 8 are marked. (a) on left, shows a nadir (ND) and glint (GL) orbit from the early mission before the spacecraft yaw offset was commanded. (b) on right, shows a nadir and glint orbit later in the mission after the spacecraft yaw offset was commanded.

2015, a 30° yaw was also adopted for nadir observations to minimize the thermal perturbations introduced when switching between nadir and glint observations [Crisp et al. 2017]. Figure 2-2b shows another pair of footprint locations, for a pair of orbits over the same ground location and a similar time of year as in Figures 2-2a. Both orbits in 2-2b occur after 12 November 2015, so they include a similar spacecraft yaw offset. Comparing the two sets of plots shows the impact of the spacecraft yaw modification described above on the ground footprint positions.

The OCO-2 glint/nadir observing strategy was refined over the first two years of operation to improve the measurement coverage and yield. The initial observing strategy recorded only glint or nadir observations over the entire sunlit hemisphere for a complete, 16-day, ground-track repeat period, and then used the other observing mode in the next 16-day cycle. This approach provided adequate coverage of oceans and continents on monthly time scales, but produced 16-day long gaps in the coverage of the ocean while in nadir mode, and limited coverage of high latitude continents while in glint mode. On 2 July 2015, this observation strategy was modified to alternate between glint and nadir observations on adjacent orbits to yield more continuous coverage of the entire sunlit hemisphere every day. On 12 November 2015, the observation strategy was optimized further to always collect glint data on orbits that are primarily over ocean, increasing the number of useful ocean observations by about 30%. All glint and nadir data collected since then used this “optimal” glint/nadir observing strategy.

The OCO-2 spacecraft can also point the instrument’s boresight at selected stationary surface targets to collect thousands of soundings in the vicinity of that site as the spacecraft flies overhead. The targeting capability is limited for OCO-2 because the coordinates of each site must be pre-programmed into the spacecraft’s computer, which can accommodate only 21 targets at one time. In addition, because it takes the OCO-2 spacecraft 2–5 minutes to acquire a target, 9–12 minutes to acquire target data, and the 2–5 minutes to return to nominal nadir or glint observations, up to a third of a dayside orbit of routine observations are lost for each target observation. With these limitations, OCO-2 acquires no more than one target observation each day. The highest priority surface targets include well characterized, vicarious calibration sites, such as Railroad Valley, Nevada, and ground validation sites, such as the Total Carbon Column Observing Network (TCCON) stations [Wunch et al. 2017]. Over the first 5.5 years of its mission, OCO-2 successfully acquired around 10 cloud-free targets each month.

### **2.2.3 OCO-3 Glint/Nadir/Target/SAM Strategy**

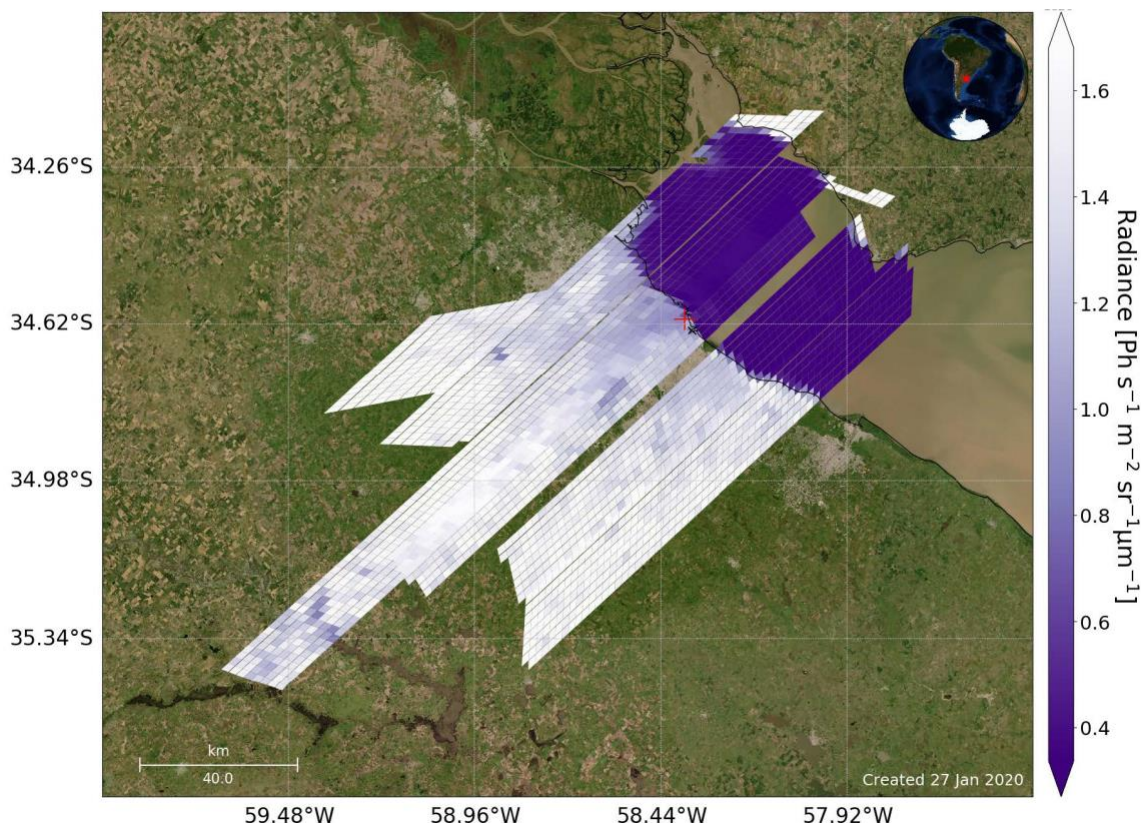
OCO-3, like OCO-2, includes a linear polarizer in front of each spectrometer slit to reject light polarized parallel to the long axis of the slit, because the gratings introduce spectrally-dependent attenuation in this component of the light. However, unlike OCO-2, the viewing geometry of the OCO-3 observations could not be constrained to a single polarization angle with respect to the principal plane. In addition, unlike the OCO-2 instrument, whose telescope observes the Earth directly, the OCO-3 telescope views the Earth through a 2-axis (altitude-azimuth) pointing mirror assembly (PMA). This PMA employs a 4-mirror design that is polarization-neutral in the absence of contamination. Because of these factors, much greater care is needed to calibrate and track the instrument’s polarization dependence (i.e. define its Mueller matrix) and the varying polarization angles in its observations to minimize polarization-dependent biases in the Xco<sub>2</sub> product. The polarization behavior of OCO-3 is described in detail in the L1B ATBD.

As noted above, OCO-3 uses its 2-axis PMA to direct the instrument boresight rather than changing spacecraft orientation. The PMA enables nadir, glint and target observations similar to those executed by OCO-2. Because the OCO-3 PMA can reorient the boresight far faster than the



OCO-2 spacecraft can, the OCO-3 observing strategy routinely switches from glint observations over ocean to nadir observations over land. While OCO-2 collects global  $X_{CO_2}$  and SIF observations as high as  $86^\circ$  from a 1:30 PM near-polar orbit, OCO-3 collects observations from the low-inclination ( $51^\circ$ ), ISS orbit. In this orbit, OCO-3 can only collect routine nadir and glint observations within  $\pm 51^\circ$  of the equator, but it can collect far more data at mid-latitudes, where the orbit transitions from ascending to descending branches. The ISS overpasses do not always occur at the same local time of day for a given point on the Earth and the ISS orbit tracks never repeat. This orbit allows OCO-3 to sample  $X_{CO_2}$  and SIF from dawn to dusk, facilitating investigations of the impact of diurnal variations in solar illumination, clouds and aerosols on the carbon cycle. These differences, and resonances between the ISS orbit and time of day yields denser measurements over the northern or southern hemispheres on an oscillating  $\sim 30$ -day cycle.

The OCO-3 instrument has a much more capable and easily programmed computer that can accommodate a much larger and much more flexible target list than the one on OCO-2. This on-board computer, combined with the fast PMA allows OCO-3 to acquire dozens of target or SAM collections each day. For target observations, the OCO-3 PMA is repositioned during the collection to do repeated scans over the target location. This process yields a typical set of 6 discrete observation geometries over a target location. Unlike the smooth pointing variation during an OCO-2 target observation. Finally, OCO-3 has a unique mode called Snapshot Area Map (SAM). Observations collected during a SAM are similar to a target mode collection, except that the PMA repointing also moves the boresight in a cross-track direction. Unlike a target mode, where the boresight is repositioned over the same ground location, the SAM



**Figure 2-3.** OCO-3 O<sub>2</sub> A-Band radiances for an OCO-3 SAM taken over Buenos Aires, Argentina at 17:22 UTC on 22 September 2019, showing the spatial coverage possible with a typical SAM data collection.



produces a raster scan over a contiguous area of up to roughly 80 km by 80 km during a single ISS overpass. Figure 2-3 shows a SAM collection from OCO-3 over Buenos Aires, Argentina. The SAM mode observations facilitate the collection of spatially-resolved maps of  $X_{CO_2}$  and SIF over targets such as large urban areas, forests, and agricultural areas.

### 2.3 Data Product Delivery

OCO-2 data are downlinked to a NASA Near Earth Network (NEN) station in Alaska during two overpasses each day. An additional downlink to the NEN station in TrollSat, Antarctica was recently made operational. From there, they are transferred to the Earth Science Mission Operations (ESMO) center at the Goddard Space Flight Center (GSFC). ESMO converts the raw telemetry into time-ordered Level 0 (L0) products and delivers these products to the OCO-2 Science Data Operations System (SDOS) at Jet Propulsion Laboratory (JPL). OCO-3 data are delivered to SDOS from the ISS Huntsville Operations Support Center (HOSC). From there, they follow the same data pipeline as the OCO-2 data.

The SDOS operators monitor and resolve disruptions in the data downlink and telemetry. SDOS also executes the processing strategy summarized in Figure 2-4 to generate Level 1A (L1A, engineering units), L1B (geolocated, calibrated radiance spectra), and L2 products (geolocated estimates of  $X_{CO_2}$ , SIF, and other atmospheric and surface state properties).

The pre-processors shown in Figure 2-4 reject soundings contaminated by clouds or have low signal-to-noise ratio (SNR) [Taylor et al. 2016]. On monthly time scales, 7–21% of all soundings pass these filters and are used to generate L2 products. In addition to the standard L2 products, the OCO-2 team produces a more compact L2 data product, called the “lite file” product. The L2  $X_{CO_2}$  and SIF products include the most commonly used L2 variables for all soundings that meet quality-screening criteria [Mandrake et al. 2015; Frankenberg 2015]. The lite files also include  $X_{CO_2}$  estimates with a global bias-correction derived by the SDOS team.

OCO-2 products are delivered as “forward” and “retrospective” processing streams. Forward stream products use extrapolated calibration coefficients and are typically delivered within one week of acquisition. The “retrospective” stream uses refined calibration coefficients and includes all cloud-free soundings processes through L2. Retrospective products generally have higher quality for science applications and are delivered at monthly intervals [Crisp et al. 2017].

The OCO-2 project archives and distributes the data products through GES DISC at <http://disc.sci.gsfc.nasa.gov/>. The OCO-3 data products are also available at GES DISC.



**Figure 2-4.** Flow diagram of data processing strategy from the raw instrument data through L2 and lite files. The data products are shown in grey, while the product generation executives (PGEs) are shown in blue.

### 3 Algorithm Description

#### 3.1 Algorithm Overview

A summary of the algorithm and its application for the v9 data of OCO-2 can be found in O'Dell et al [2018].

The primary purpose of the retrieval algorithm is to derive estimates of the column-averaged atmospheric CO<sub>2</sub> dry air mole fraction,  $X_{CO_2}$ , and other L2 data products from the spectra returned by the OCO-2 mission.  $X_{CO_2}$  is defined as the ratio of the column abundances of CO<sub>2</sub> to the column abundance of dry air:

$$X_{CO_2} = \frac{\int_0^\infty N_{CO_2}(z)dz}{\int_0^\infty N_{air}(z)dz} \quad (3-1)$$

Here,  $N_{CO_2}(z)$  is the altitude ( $z$ )-dependent number density of CO<sub>2</sub> (e.g., number of CO<sub>2</sub> molecules per cubic meter) and  $N_{air}(z)$  is the altitude-dependent number density of dry air. Because O<sub>2</sub> constitutes 0.20935  $N_{air}$ ,  $X_{CO_2}$  can also be expressed as:

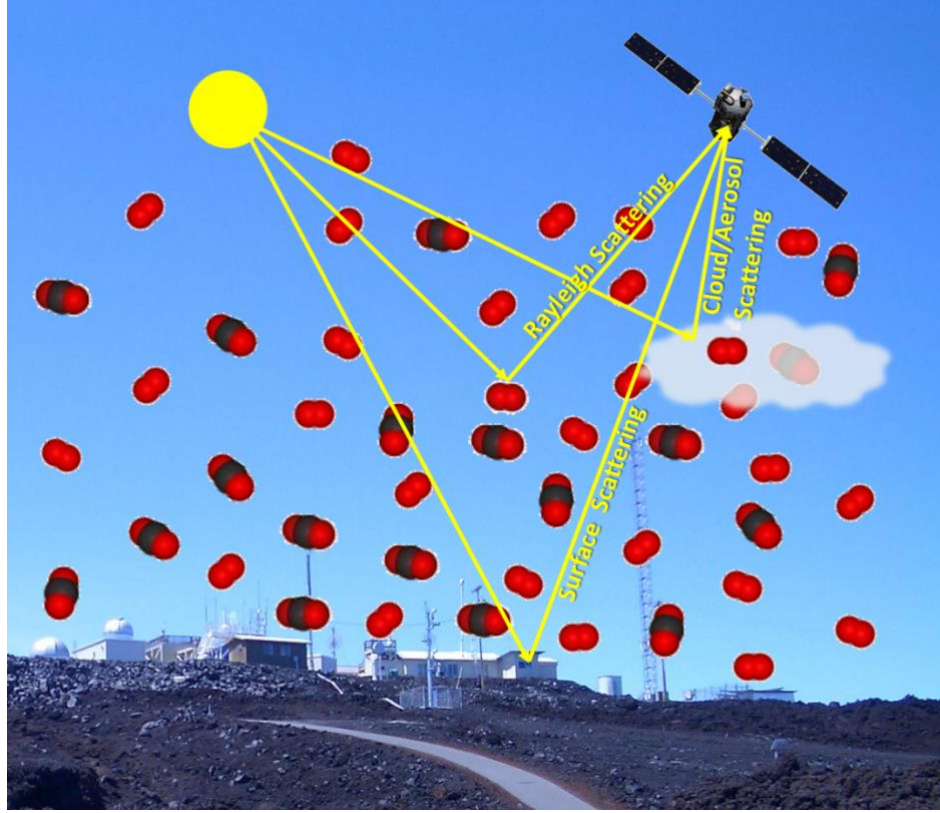
$$X_{CO_2} = \frac{0.20935 \cdot \int_0^\infty N_{CO_2}(z)dz}{\int_0^\infty N_{O_2}(z)dz} \quad (3-2)$$

The number densities of CO<sub>2</sub> and O<sub>2</sub> can be inferred from precise spectroscopic observations of reflected sunlight because the measured intensity of the sunlight at wavelengths where these gases absorb is proportional to the total number of molecules along the optical path  $S$ :

$$I(\lambda, \theta, \theta_0, \varphi - \varphi_0) = F_0(\lambda) \cos \theta_0 \cdot R(\lambda, \theta, \theta_0, \varphi - \varphi_0) \left\langle \exp \left\{ - \int_0^S \sum_{m=1}^M [\sigma_m(\lambda, s) N_m(s)] ds \right\} \right\rangle \quad (3-3)$$

Here,  $I(\lambda, \theta, \theta_0, \varphi - \varphi_0)$  is the observed intensity at wavelength  $\lambda$ ,  $\theta$  and  $\varphi$  are the observation zenith and azimuth angles, and  $\theta_0$  and  $\varphi_0$  are the corresponding solar zenith and azimuth angles.  $F_0(\lambda)$  is the solar flux at the top of the atmosphere,  $R(\lambda, \theta, \theta_0, \varphi - \varphi_0)$  is the reflectance of the surface,  $\sigma_m(\lambda, S)$  and  $N_m(S)$  are the absorption cross sections of the  $m^{th}$  absorbing constituent (i.e., CO<sub>2</sub>, O<sub>2</sub>, or other absorber), and the integration is performed along an optical path  $S$ , which extends from the top of the atmosphere to the reflecting surface and back to the spacecraft. The brackets  $\langle \rangle$  indicate an average over the range of possible optical paths that the photons could travel. A few of the possible optical paths are illustrated in Figure 3-1.

Because the argument of the exponential in Equation (3-3) depends linearly on the optical cross section per molecule,  $\sigma$ , and length of the optical path,  $S$ , as well as the number density of absorbing molecules, errors or uncertainties in either  $\sigma$  or  $S$  can introduce compensating errors and uncertainties in the retrieved number densities. For example, consider a sounding footprint that is partially obscured by a cloud, which scatters a fraction of the radiation back to the sensor before it has traversed the complete path from the top of the atmosphere to the surface and back to the spacecraft. If the scattering by this cloud is neglected in the retrieval algorithm, the



**Figure 3-1.** Viewing geometry and a few possible optical paths traversed by solar photons that are recorded by the instrument.

reduced optical path could be misinterpreted as evidence for a reduced number density of absorbing molecules. Similarly, if multiple scattering between the surface and an airborne aerosol layer produces an optical path that is longer than the direct path between the Sun, surface, and spacecraft, and this multiple scattering is neglected, the enhanced optical path could be misinterpreted as evidence for an increased number density of absorbing molecules. Accurately estimating the atmospheric optical path is therefore a primary focus of the retrieval algorithm.

If we divide the atmosphere into a series of discrete layers where the vertical distributions of the absorbing constituents are uniform or at least spatially-uncorrelated, we can rewrite the quantity:

$$\int_0^s \sum_{m=1}^M [\sigma_m(\lambda, s) N_m(s)] ds = \sum_{n=1}^N \sum_{m=1}^M \int_{s(n)}^{s(n+1)} [\sigma_m(\lambda, s) N_m(s)] ds \quad (3-4)$$

If we consider a linear path through a plane-parallel layer  $n$  at a local zenith angle of  $\theta$ , the optical path between point  $s(n)$  and  $s(n+1)$  can be approximated as  $ds = dz/\cos\theta$ . We can then express the quantity included in the integrand in terms of the vertical optical depth in each layer:

$$\tau_{m,n}(\lambda, z) = \int_{s(n)}^{s(n+1)} [\sigma_m(\lambda, z) N_m(z)] dz \quad (3-5)$$

The total column-integrated vertical optical depth can then be given as:

$$\tau(\lambda) = \sum_{n=1}^N \sum_{m=1}^M \int_{s(n)}^{s(n+1)} [\sigma_m(\lambda, z) N_m(z)] dz \quad (3-6)$$

If we assume that the atmosphere is plane-parallel and ignore scattering by gases and airborne particles, the intensity observed by the spacecraft for a solar zenith angle  $\theta_0$  and observation angle  $\theta$  is given by:

$$I(\lambda, \theta, \theta_0, \varphi - \varphi_0) = F_0(\lambda) \cos \theta_0 \cdot R(\lambda, \theta, \theta_0, \varphi - \varphi_0) \times \exp \left\{ -\tau(\lambda) \frac{\cos \theta_0 + \cos \theta}{\cos \theta_0 \cdot \cos \theta} \right\} \quad (3-7)$$

In the above discussion, the complexity associated with scattering by gases, airborne particles, and the surface was avoided by introducing the optical path averaging operator  $\langle \rangle$  or was simply ignored (as in Equation (3-6)). More generally, the observed intensity will depend on both absorption and scattering, which can alter the optical path traversed by the solar photons. Fortunately, all of these optical processes can be accurately simulated by solving the equation of radiative transfer:

$$\mu \frac{d\mathbf{I}(\tau, \theta, \phi)}{d\tau} = \mathbf{I}(\tau, \theta, \phi) - \mathbf{J}(\tau, \theta, \phi) \quad (3-8)$$

where

$$\mathbf{J}(\tau, \theta, \phi) = \frac{\omega(\tau)}{4\pi} \int_{-1}^1 \int_0^{2\pi} \mathbf{\Pi}(\tau, \theta, \theta', \phi - \phi') \mathbf{I}(\tau, \theta', \phi') d\phi' d\theta' + \mathbf{Q}(\tau, \mu, \phi) \quad (3-9)$$

Here,  $\omega$  is the single scattering albedo,  $\mathbf{\Pi}$  is the phase matrix for scattering, and  $\mu = \cos \theta$ . The first term in Equation (3-9) represents multiple scattering contributions. If we ignore thermal emission (an accurate assumption over the range of wavelengths observed by OCO-2), the inhomogeneous source term  $\mathbf{Q}(\tau, \mu, \varphi_0)$  describing single scattering of the (attenuated) solar beam can be expressed as:

$$\mathbf{Q}(\tau, \mu, \varphi - \varphi_0) = \frac{\omega(\tau)}{4\pi} \mathbf{\Pi}(\tau, \mu - \mu_0, \phi - \phi_0) \mathbf{I}_0 \exp[-\tau / \mu_0] \quad (3-10)$$

Within the context of an atmospheric remote-sensing retrieval algorithm, the equation of radiative transfer is a component of the forward model.

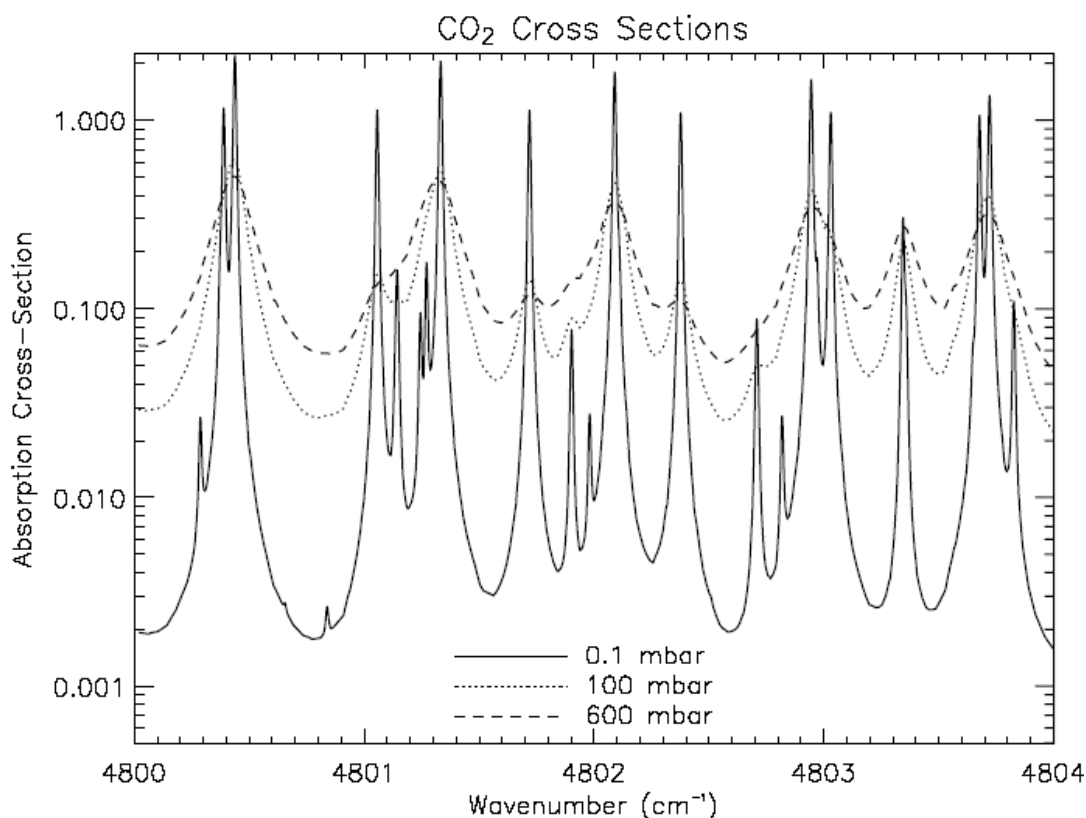
Monitoring small changes in the abundances of gases such as O<sub>2</sub> and CO<sub>2</sub> using NIR absorption poses special challenges at NIR wavelengths because their absorption cross sections vary rapidly with wavelength (Figure 3-2).

The strongest absorption occurs near the centers of the narrow vibration-rotation lines. At other wavelengths, the absorption can be orders of magnitude weaker. Hence, for small number densities and/or optical path lengths, changes in these properties will produce their largest

variations in the radiance field at wavelengths near the line centers. Eventually, however, as the number densities or optical path lengths increase, virtually all of the sunlight is absorbed at these wavelengths (e.g., the line cores become saturated), and the sensitivity to further changes decreases dramatically. The region of peak sensitivity then moves further down the line wings.

Even with its relatively high spectral resolving power ( $\lambda/\delta\lambda \sim 20,000$ ), the OCO-2 instrument does not completely resolve the narrow cores of spectral lines. In spite of this, the spectral variability must be resolved by the forward model to accurately quantify the absorption within the spectral range sampled by the instrument line shape (ILS) function of each individual detector pixel. The most straightforward way to do this is to divide the spectral range of interest into a series of discrete spectral intervals that are narrow enough to completely resolve the spectral features contributed by atmospheric gases, the incident solar radiation field, airborne particles, or the reflecting surface. The equation of radiative transfer must then be evaluated within each spectral interval. Finally, the spectrum observed by each detector pixel is the product of the full-resolution simulated spectrum and the ILS function for the pixel.

The primary advantages of this direct spectrum-resolving approach are its conceptual simplicity and intrinsic accuracy. It can fully accommodate all of the physics of the atmospheric and surface processes that contribute to the absorption and scattering of solar radiation. Its primary drawback is its computational expense, because the equation of radiative transfer must be evaluated thousands of times to resolve the spectral structure within an NIR O<sub>2</sub> or CO<sub>2</sub> band. A number of methods (correlated-k, spectral mapping, etc.) have been developed to reduce the number of spectral intervals needed, but all of these methods employ simplifications and approximations that can introduce unacceptable errors in the computed radiance spectra.



**Figure 3-2.** Spectral dependence of the absorption coefficients of CO<sub>2</sub> at wavelengths within the strong CO<sub>2</sub> band near 2.08  $\mu\text{m}$  are shown for pressures near 0.1, 100, and 600 hPa (mbar).

To retrieve estimates of  $N_{CO_2}(z)$  or  $N_{O_2}(z)$  (or  $\tau_{m,n}$ ) from observations of reflected solar radiation, the equation of transfer must be inverted. In general, this cannot be done analytically because of the complexity and nonlinearity of the equation of radiative transfer. However, these quantities can be derived using conventional nonlinear least-squares fitting techniques, where these quantities are treated as unknown coefficients and the equation of radiative transfer is the fitting function. In addition to the fitting function, these techniques require the first derivatives of the intensities with respect to any specific component of the state vector  $x_{m,n}$ :

$$K_{m,n} = \frac{\partial I}{\partial x_{m,n}} \quad (3-11)$$

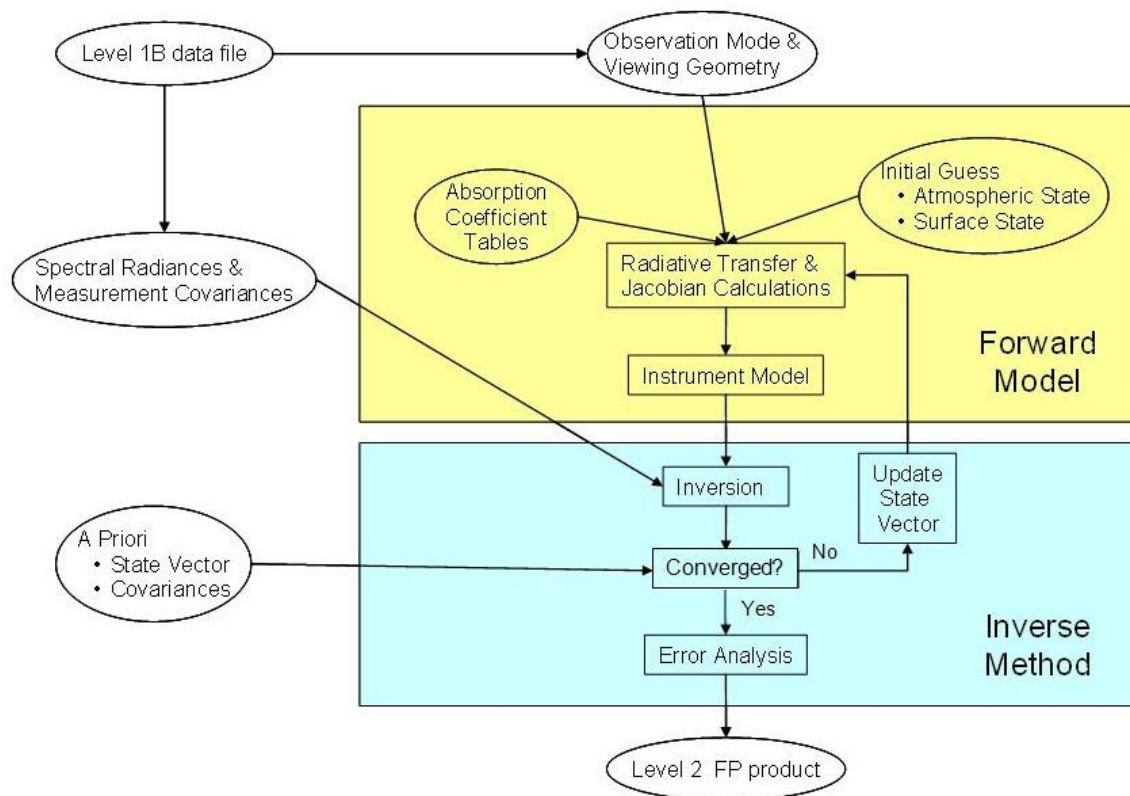
These radiance Jacobians are also generated by the forward model. Because the equation of radiative transfer is highly nonlinear with respect to most elements of the state vector, radiances and Jacobians must be derived for a surface/atmospheric state that is close to the actual state or the retrieval algorithm may not converge.

The primary components of the OCO-2  $X_{CO_2}$  retrieval process are summarized in Figure 3-3. Given an initial guess for the atmospheric and surface states and the observing geometry for a specific sounding, the forward model generates polarized synthetic radiance spectra and radiance Jacobians in the  $O_2$  A band and in the two  $CO_2$  bands observed by the OCO-2 instrument. This model first generates these synthetic spectra on a spectral grid that fully resolves the solar spectrum at the top of the atmosphere, the absorption and scattering cross sections for each atmospheric gas and airborne cloud and aerosol particle type included in the state vector, and the reflecting surface. The instrument model then convolves this full-resolution synthetic radiance spectrum using the ILS function and corrects for instrument polarization to simulate the spectrum recorded by the instrument. The forward model is described in Section 3.2.

The state structure includes atmospheric and surface properties that affect the spectrally-dependent radiances observed by the satellite. The surface-atmosphere state vector includes those components of the state structure that will be optimized by the retrieval algorithm to improve the fit between the observed and simulated radiances. The state vector is described in Section 3.3. Some of aspects of the instrument throughput (e.g., dispersion, instrument line shape) can also be included in the state vector. The methods used to generate radiance Jacobians are described in Section 3.4.

The inverse method is based on a Rodgers [2000]-type optimal estimation approach and has been described in Connor et al. [2008] and Boesch et al. [2006]. This model modifies the initial state vector to minimize differences between the observed and simulated spectra from each sounding. The inverse method is described in Section 3.5.

Once the atmospheric state yielding the best match to the observed spectrum has been found, the algorithm then determines  $X_{CO_2}$ , errors in  $X_{CO_2}$  from different sources (such as vertical smoothing, measurement noise, etc.), and the  $X_{CO_2}$  column averaging kernel. This step is necessary because  $X_{CO_2}$  is not itself an element of the state vector. Rather, it is determined from the profile of  $CO_2$ , which is part of the state vector. It is formally given by the total number of  $CO_2$  molecules in the column divided by the total number of dry air molecules in the column. This step is labeled “Error Analysis” in Figure 3-3 and is described in Section 3.6.



**Figure 3-3.** Flow diagram for the L2 retrieval algorithm.

Finally, given an estimate of  $X_{CO_2}$ , a series of screening tests are performed to evaluate the quality of the retrieval. The criteria for a successful retrieval are currently under investigation and are subject to change prior to the final delivery of the data. Examples of success criteria that have been used in the past are:

- Measurement relative  $\chi^2$  in each band < threshold
- *a posteriori*  $X_{CO_2}$  error < threshold
- Retrieved aerosol optical depth < threshold

A new method to assign warn levels to the data has been developed, as described in Mandrake et al. [2013]. This approach uses an objective method to bin the data into 20 groups and assign a warn level, with the lowest numbers indicating the best data quality, and increasing warn levels as data quality decreases. These warn levels will be included with the data as an alternative to the empirical data screening described above.

### 3.2 Forward Model

The forward model simulates solar spectra and radiance Jacobians for analysis of observations acquired by the OCO-2 spacecraft in nadir, glint, and target modes. A simple cartoon illustrating the atmospheric and surface optical processes that contribute to the absorption and scattering of solar radiation as it traverses the atmosphere and is reflected by the surface is shown in Figure 3-1. The forward model can also generate synthetic spectra and Jacobians for the analysis of direct observations of the solar disk collected by the OCO-2 instrument during thermal vacuum testing or those collected by the Fourier transform spectrometers in the TCCON.



The software components of the forward model include:

- Spectrally-dependent surface and atmosphere optical properties
  - Cloud and aerosol single scattering optical properties
  - Gas absorption and scattering cross-sections
  - Surface reflectance
- Spectrum effects
  - Solar model
  - Fluorescence
  - Instrument Doppler shift
- Atmosphere/surface radiative transfer model
- Instrument model
  - Spectral dispersion
  - ILS
  - Polarization response
  - Zero-level offset (optionally)

The methods used to generate the spectrally dependent gas and particle optical properties are described in Section 3.2.1. The solar spectrum at the top of the atmosphere is generated by a model described in Section 3.2.2. Synthetic spectra are generated by a radiative transfer algorithm based on a spectrum-resolving, multi-stream multiple scattering radiative transfer model that incorporates the linear discrete ordinate radiative transfer model (LIDORT) [Spurr et al., 2001; Spurr, 2002], a polarization correction based on a fast two-orders-of-scattering model [Natraj and Spurr, 2007], and a series of other techniques to improve its accuracy, speed, and range of validity. This model is described in Section 3.2.3. The instrument model simulates the spectral dispersion, ILS function, polarization dependence, and any other necessary corrections, such as zero-level offset. The instrument model is described in Section 3.2.4. The inputs and outputs of the forward model are enumerated in Table 3-1.

Several assumptions have been made for analyzing the OCO-2 data. First, it is assumed that measured radiances have been radiometrically calibrated. Second, it is assumed that thermal emission from the surface and atmosphere is negligible compared with reflected sunlight. This is certainly true in the O<sub>2</sub> A and weak CO<sub>2</sub> bands, and is a reasonably good assumption in the strong CO<sub>2</sub> band at 2.06 microns. Third, soundings that contain optically-thick clouds or aerosols can be rejected because they contain no useful information about CO<sub>2</sub> in the lower troposphere. It is assumed that inelastic scattering processes (e.g., Raman) are negligible at the wavelengths of interest to OCO-2. Further, the effects of airglow and absorption by the Chappuis bands of ozone in the O<sub>2</sub> A band are also currently ignored.

**Table 3-1.** Inputs and outputs of the forward model.

Inputs	Outputs
State structure: <ul style="list-style-type: none"> <li>• Atmospheric state</li> <li>• Surface state</li> <li>• Instrument (ILS, dispersion)</li> </ul>	Radiance spectrum



Inputs	Outputs
Observation geometry: zenith angles, altitude, etc.	Partial derivatives (Jacobians) of the radiance spectrum with respect to each of the state vector elements
Solar absorption and continuum tables	
Gas absorption and scattering cross sections	
Aerosol optical properties	
Surface bidirectional reflectance distribution function parameters	

### 3.2.1 Optical Properties

This section describes the atmospheric absorption and scattering calculation in the forward model. The radiative transfer model uses these values to predict the solar energy absorbed or scattered at each wavelength within an atmospheric layer at a given temperature and pressure. To improve computational efficiency, lookup tables for the absorption cross-sections for each absorbing gas are computed in advance. The OCO-2 forward model interpolates between table entries at runtime. For absorbing gases, the table generation process records representative cross sections  $k_\nu$  ( $\text{cm}^2/\text{molecule}$ ) at *every* applicable wavelength for the full range of applicable pressures (Pa), temperatures (K), frequencies ( $\text{cm}^{-1}$ ), and broadeners such as water vapor. The resulting absorption coefficients (ABSCO) table is a four-dimensional lookup table for each absorbing gas in the forward model. For airborne cloud and aerosol particles, the wavelength single-scattering optical cross sections and scattering phase functions are tabulated for specific particle modes. The wavelength-dependent Rayleigh scattering cross sections and phase functions are calculated for each iteration for each retrieval.

We begin with a broad overview of the gas absorption calculations and physical models used to produce the cross sections, followed by a brief review of the laboratory experiments and spectroscopic parameters used to inform these calculations. We then describe software that implements the algorithm.

#### 3.2.1.1 Gas Absorption Cross Section Calculation

The OCO-2 L2 algorithm uses precalculated lookup tables of absorption coefficients (ABSCO tables) for the calculation of gas absorption cross sections. The ABSCO tables contain molecular absorption cross sections over the range of relevant wavelengths, temperatures, pressures, and broadeners in units of  $\text{cm}^2/\text{molecule}$ . The forward model in the L2 algorithm computes atmospheric absorption at each relevant temperature, pressure, and wavelength using linear interpolation. New laboratory results and theoretical models have been incorporated into successive versions of ABSCO tables in order to obtain more accurate absorption coefficients. ABSCO tables are currently available for  $\text{O}_2$ ,  $\text{CO}_2$ ,  $\text{H}_2\text{O}$ , and  $\text{CH}_4$ , and are released intermittently.

For a single molecule, the absorption associated with a single transition  $k$  between two states  $i$  and  $j$  centered at wavenumber  $\nu_k (= \nu_{ij})$  can be written as the product of the corresponding line intensity  $S_k (= S_{ij})$  and a line shape function  $f(\nu, \nu_k)$ . The monochromatic absorption coefficient at frequency  $\nu_k$  of a given species is given by a summation over all relevant lines:

$$k_\nu^x = \sum_{\nu} S_k f(\nu, \nu_k) \quad (3-12)$$

The line shape function most commonly used within codes for the calculation of atmospheric absorption is the Voigt line shape. This shape is a convolution between the Lorentz (or pressure-broadening) function, which is applicable to higher pressures where the shape of the spectral lines is dominated by collisions between molecules, and the Doppler line shape, which describes a low-pressure regime in which the line shape is determined by the distribution of molecular velocities.

The Voigt line shape offers a number of advantages in terms of the availability of fast algorithms and the availability of line parameters. For example, the High Resolution Transmission (HITRAN) 2012 database [Rothman et al., 2013] provides Voigt line shape parameters for 47 different molecules, including over 470,000 lines for CO<sub>2</sub> alone. However, a number of studies have shown that the Voigt line shape is not sufficient for accurate retrievals of CO<sub>2</sub> and O<sub>2</sub> from near-IR solar spectra at the level required to meet OCO-2 science objectives (e.g., Long and Hodges [2012], Tran and Hartmann [2008], Predoi-Cross et al. [2008a], and Hartmann et al. [2009]). The OCO-2 retrievals require consideration of more subtle phenomena such as line mixing, speed dependence, and collision-induced absorption.

Line mixing occurs when the transition energies of a given molecular species overlap such that rotationally-inelastic collisions can transfer intensity from one line to another. Line mixing can change the population of states (i.e., the distribution of line strengths within a band) as well as the shape of individual spectral lines, overestimating the absorption in the wings of the band and underestimating the absorption in the peaks of the P and R branches [Tran et al., 2006]. Speed dependence of molecular collisions leads to deviation from the basic Voigt line shape (where all collisions are assumed to occur at the peak of the velocity distribution). Collision-induced absorption (CIA) occurs at high pressures, where interactions between molecules are frequent enough to allow observation of unbound collisional-complexes. The collisional complexes with high enough density for observation are those composed of dioxygen pairs and dioxygen-dinitrogen molecules, and these complexes naturally exhibit pre-dissociated spectra in the same spectral regions as the monomers, resulting in non-negligible contributions to opacity in the O<sub>2</sub> A-band region.

Line mixing, speed dependence, and CIA are all currently considered within the OCO-2 ABSCO model. However, line shape is still an active area of research and there is still significant room for improvement in the modeling of absorption coefficients in the OCO-2 bands. Further details on specific molecules are provided below.

### **3.2.1.2 Spectroscopy for OCO-2 Bands**

In order to achieve improvements to the near-IR spectroscopy of CO<sub>2</sub> and O<sub>2</sub> in the three spectral channels at 0.76, 1.61, and 2.06  $\mu\text{m}$ , members of the OCO-2 Science Team have formed international collaborations with leading experts to obtain high-quality laboratory measurements and theoretical modeling of high-resolution laboratory spectra of these two constituents.

In the last ten years, the state of the art has greatly advanced in measuring the spectral line positions, line intensities, and pressure-broadened line shapes and pressure shifts of CO<sub>2</sub> and O<sub>2</sub> absorption features (e.g., Miller et al. [2004a, b; 2005], Toth et al. [2006a, b; 2007a, b; 2008a, b], Devi et al., [2007a, b], Predoi-Cross et al. [2009], Sung et al. [2009], Devi et al. [2016], Benner et al. [2016], and Long et al. 2020 for CO<sub>2</sub>, and Robichaud et al. [2008a, b], Robichaud et al. [2009], Predoi-Cross [2008a, b], Tran and Hartmann [2008], and Drouin et al. [2017] for O<sub>2</sub>). Improvements came from careful control of the experimental process applied for the laboratory measurements, use of reliable spectrometers (the McMath-Pierce Fourier transform spectrometer at Kitt Peak National Observatory, a Bruker 125 FTS at JPL, a Cavity ring-down spectrometer

[CRDS] at National Institute of Standards and Technology [NIST] in Maryland, a Bomem FTS at the National Research Council in Canada, a Bruker 120 FTS in Giessen, Germany, and a Bruker FTS 66V at Laboratoire Interuniversitaire des Systèmes Atmosphériques in France), and sophisticated retrieval methods that evoked non-Voigt line shapes in a multispectrum fitting procedure (e.g., Devi et al. [2007a, b]).

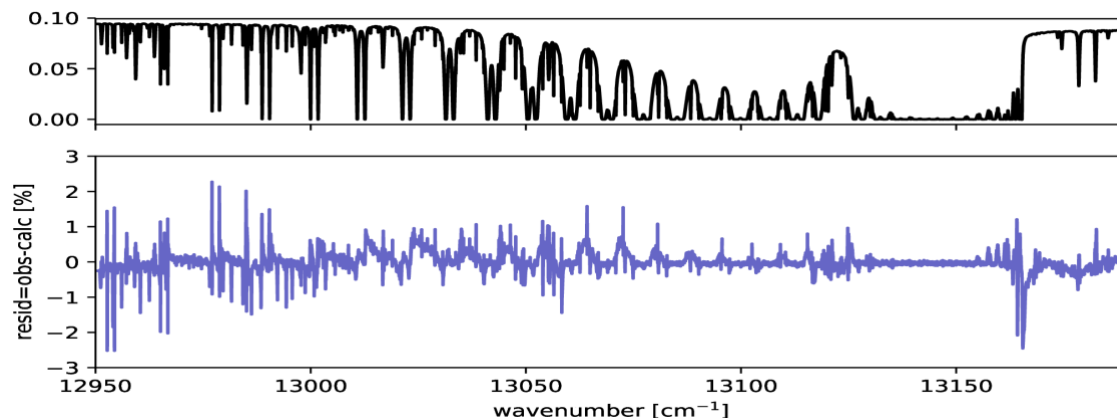
#### **3.2.1.2.1 Oxygen A Band**

Versions of ABSCO prior to v5.0 relied on code developed by Tran et al. [2006] for the main isotopologue of O<sub>2</sub>. Since then, the LABFIT codebase developed by Benner et al. [2011] has been extended to accommodate the parameterization of O<sub>2</sub> from laboratory spectra. To include the effects of line mixing and CIA, we have fitted laboratory spectra, supplemented internally and externally with atmospheric data to best couch the effects measured in the laboratory spectra that must be extrapolated to atmospheric conditions. The multispectrum fitting simultaneously solved for lineshape parameters of a speed-dependent Voigt for strong lines while also determining band-wide parameters describing line position and intensity [Drouin et al., 2017; Payne et al., 2020]. We supplemented them with a few additional Voigt-profile lines from the HITRAN 2012 database to cover all remaining weak O<sub>2</sub> lines in the entire 12850 – 12900 cm<sup>-1</sup> window. An ad hoc line-mixing model was adapted from Tran et al [2006] but did not include the ad hoc empirical corrections applied by Tran and Hartman [2008]. Adjustments to a laboratory based CIA model were made in concert with changes to the line-mixing model. The model that best fit FTS and CRDS laboratory spectra was tested with TCCON data [Wunch et al., 2010] and an atmospheric based CIA was determined from these tests for inclusion in the ABSCO tables.

Broadening of O<sub>2</sub> by H<sub>2</sub>O was measured in the 60 GHz Q-branch (50 – 119 GHz) and in the pure rotational S-branch (424 – 1850 GHz) by Drouin et al. [2014]. We assume no vibrational dependence and apply the Drouin et al. [2014]. H<sub>2</sub>O broadening coefficients to the O<sub>2</sub> A band also.

Tables generated using the parameters and software detailed above have been validated using spectra from independent laboratory measurements and ground-based FTS measurements from the TCCON network and the Greenhouse Gases Observing Satellite (GOSAT). These comparisons show systematic residuals above the level of instrument noise, indicating significant room for improvement in modeling of O<sub>2</sub> absorption. Example residuals are shown in Figure 3-4.

The L2 algorithm contains the capability to apply empirical scaling factors to the ABSCO tables. For OCO-2 B10 retrievals, an empirical scaling of 1.0048 has been applied to ABSCO v5.1 to reduce biases in retrieved surface pressure. Example residuals in the OCO-2 O<sub>2</sub>A-band are shown in Figure 3-4.



**Figure 3-4.** Example mean spectral residual for high airmass ( $\text{sza} > 75^\circ$ ) for the O2A band using the same set of TCCON spectra used in the analysis of the absco 5.1 update for CO<sub>2</sub> (Oyafuso [2017]). Spectroscopic residuals are evident near the band head, between the P and R branch and in between doublets in the P branch.

Work continues to incorporate new laboratory measurements into a multispectrum fit for the O<sub>2</sub> A band. This approach will allow us to capitalize on the strengths of a range of different complementary laboratory measurement techniques. High-quality CRDS measurements (from NIST) offer extremely high precision and the potential of better constraints on the line shape (including line mixing and CIA effects). FTS measurements can ensure consistency of measurement across wide spectral ranges. New FTS measurements (taken at JPL) offer a wealth of potential new information on line shape, line mixing, and temperature dependence. Photoacoustic spectroscopy measurements taken at Caltech may offer information at high pressure, providing new insights into CIA.

#### **3.2.1.2.2 Carbon Dioxide Bands**

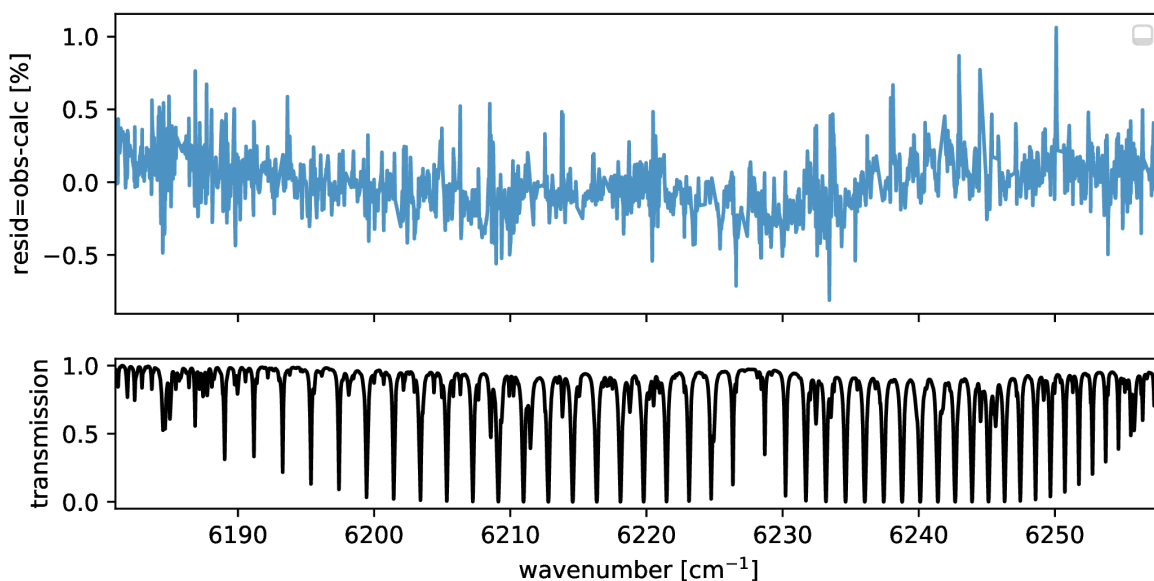
Prior to ABSCO v5.0, the CO<sub>2</sub> bands used line parameters and mixing models derived from studies by Benner et al. [2011] and Devi et al. [2007a]. ABSCO v5.1 (whose CO<sub>2</sub> cross sections are unchanged from those of ABSCO v5.0) use updates detailed in Devi et al. [2016] and Benner et al. [2016]. Additional details on validation experiments for these databases using TCCON observations are described in Oyafuso [2017]. The computation incorporates a speed-dependent line profile with tridiagonal line mixing. These databases were found to result in improved spectral residuals for atmospheric measurements compared to the databases generated using first-order line mixing, computed using software from Niro et al. [2005a, b]. Parameters for foreign broadening of CO<sub>2</sub> by H<sub>2</sub>O come from studies by Sung et al. [2009]. This Sung et al. study treated the 4.3  $\mu\text{m}$  band only. However, it is likely that these H<sub>2</sub>O broadening parameters can be safely extrapolated to the other bands as well, and we favor this approach until additional direct laboratory data is available.

Versions of ABSCO prior to v5.0 found an improvement in fits to measurements from the TCCON network of ground-based, upward-looking FTS stations by making slight adjustments to the relative HITRAN isotopologue abundances [Thompson et al., 2012]. Fractionation effects due to varying molecular weights can lead, in principle, to measurably different relative abundances among the isotopologues. However, tests with TCCON data using the latest absorption coefficients found no significant improvement in residuals in allowing the relative isotopologue abundances to depart from HITRAN values. Therefore, ABSCO 5.1 uses relative isotopologue abundances unmodified from those of HITRAN 2012.

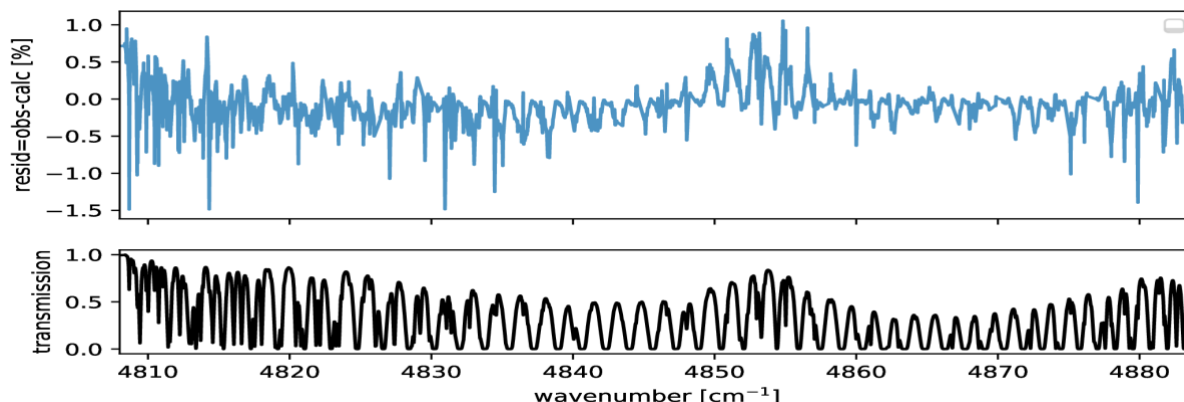
The necessity for a contribution of CO<sub>2</sub> “CIA” underlying the allowed ro-vibrational transitions in the 2.06  $\mu\text{m}$  CO<sub>2</sub> band remains. As discussed in Thompson et al. [2012], we found it necessary to incorporate an additional “continuum” absorption in the 2.06  $\mu\text{m}$  CO<sub>2</sub> band, which takes the shape of two Gaussian distributions centered at 4853.5 and 4789 wavenumbers, with standard deviations of 10 and 8 wavenumbers and intensity scalings of  $2.1 \times 10^{-24}$  and  $4.2 \times 10^{-25}$ , respectively (this is the intensity at 1 Atm, but absorption at other levels scales in proportion to pressure). These parameters were set to minimize error in retrievals with TCCON up-looking FTS spectra. Tests done in advance of future ABSCO releases have shown that a more complete line-mixing module would negate the need for these empirically-based spectral features in the ABSCO tables.

A comparison of line intensities in the 1.6  $\mu\text{m}$  band determined in Devi et al. [2016] found systematic differences from reference intensity measurements from the National Institute for Standards and Technology described in Polyansky [2015]. To align the line intensities, a scaling factor of 1.014 is applied to the cross sections prior to ingestion by the OCO-2 retrieval software. No prior scaling is applied to the 2.06  $\mu\text{m}$  band, because reference intensities were not available when the cross section tables were created. However, the L2 algorithm has the capability to apply additional ad hoc scaling factors to the delivered ABSCO tables. Currently, the L2 algorithm applies no additional scaling factor to the 1.6  $\mu\text{m}$  band but does rescale the 2.06  $\mu\text{m}$  band by a factor of 1.006.

Example residuals in the OCO-2 CO<sub>2</sub> bands are shown in Figure 3-5 and Figure 3-6. The current ABSCO tables offer improvements over previous versions, but the resulting residuals do show evidence of persistent systematic residuals, particularly in the 2.06  $\mu\text{m}$  band. Future work will involve utilization of new low temperature laboratory FTS measurements (from JPL) and precision intensity CRDS measurements (from NIST) in the multispectrum fits for CO<sub>2</sub>, in order to provide improved constraints on temperature dependence and line shape.



**Figure 3-5.** Residuals for high airmass ( $\text{sza} > 75$  degrees) atmospheric spectra at 1.6  $\mu\text{m}$ . Data set consists of 164 spectra and several problematic water lines have been masked out as described in Oyafuso et al. [2017].



**Figure 3-6.** Residuals for high airmass atmospheric spectra ( $\text{sza} > 75$  degrees) at  $2.06 \mu\text{m}$ . See text associated with Error! Reference source not found.. Figure adapted from Oyafuso et al. [2017].

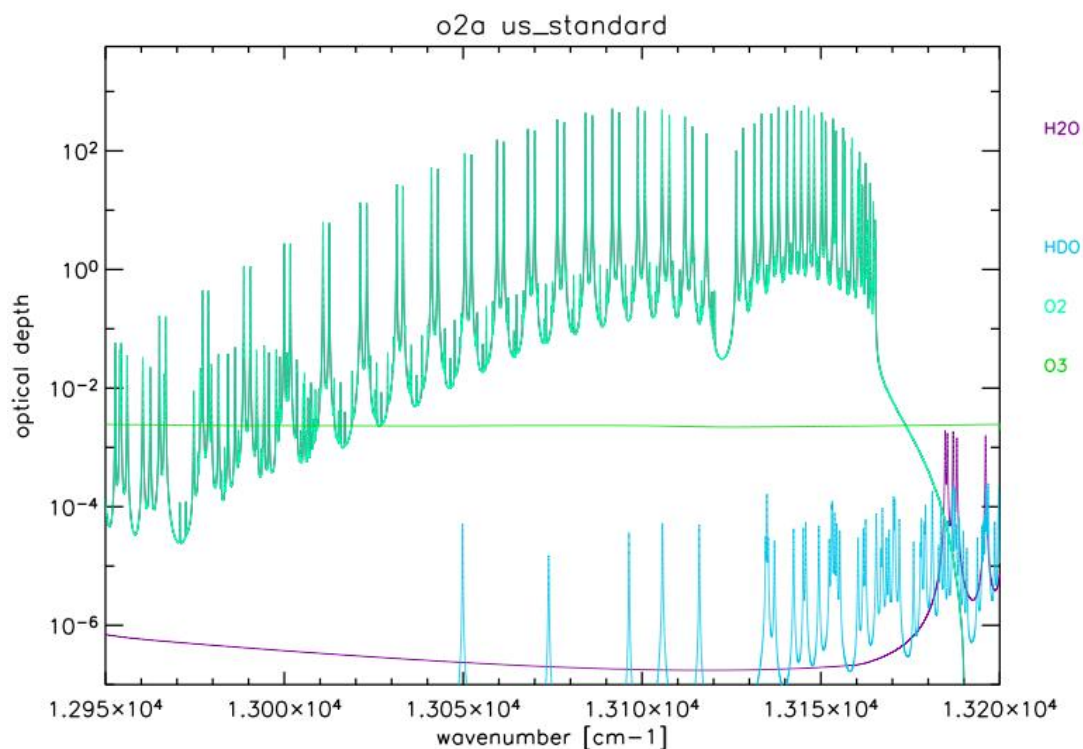
### 3.2.1.2.3 Other Absorbing Gases

Several gases besides  $\text{CO}_2$  and  $\text{O}_2$  absorb within the spectral ranges used by OCO-2. Optical depth contributions for these gases (calculated using the Line by Line Radiative Transfer Model available at <http://rtweb.aer.com>) are shown in Figure 3-7, Figure 3-8, and Figure 3-9.

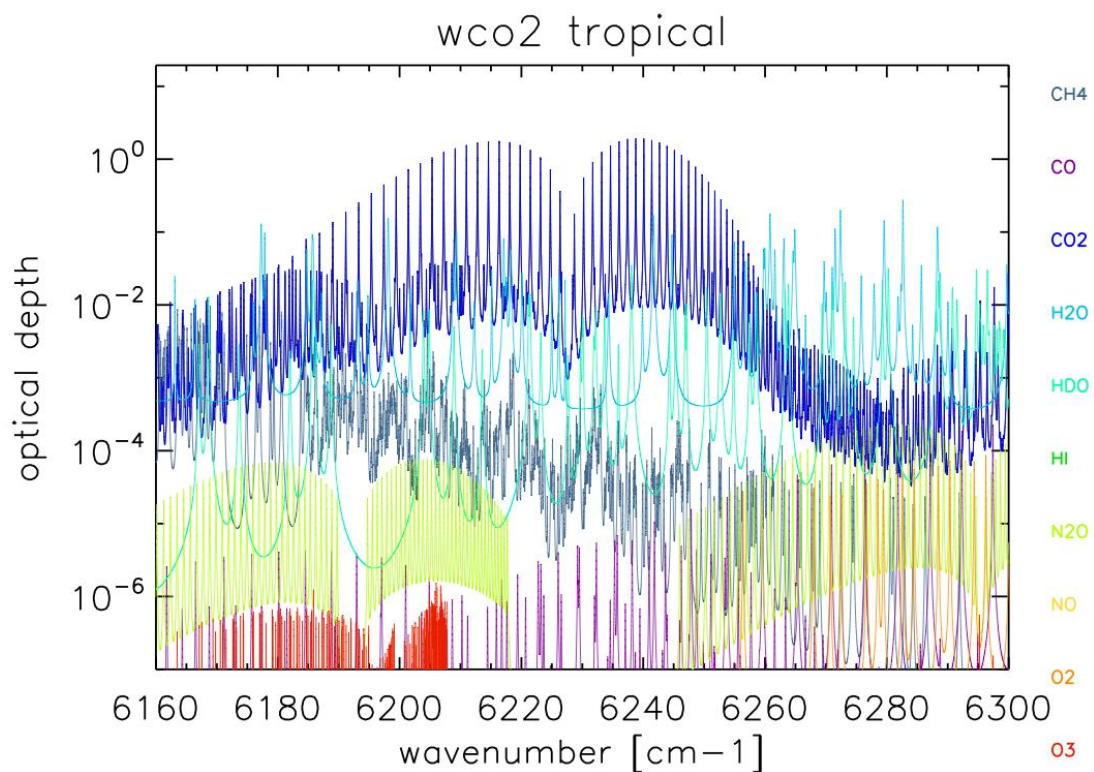
The most important gas is water vapor ( $\text{H}_2\text{O}$ ), which produces measurable absorption in all three bands. Line parameters for  $\text{H}_2\text{O}$  transitions in the OCO-2 channels are currently taken from the 2012 version of HITRAN. HITRAN 2016 was evaluated for ABSCO 5.1, and despite showing general overall improvement over a large spectral range, it did not offer an unambiguous improvement in residuals over the specific spectral ranges used by OCO-2. Therefore, HITRAN 2012 was retained in ABSCO 5.1. The  $\text{H}_2\text{O}$  continuum model used in the OCO-2 ABSCO tables is the MT\_CKD model version 3.2 with updates as described in a 2014 OCO-2 Science Team Meeting presentation by Mlawer et al. [2012,2014].

The second most important interferent is methane ( $\text{CH}_4$ ). Absorption coefficients for  $\text{CH}_4$  have been calculated using the HITRAN 2008 database [Rothman et al., 2009]. Again,  $\text{CH}_4$  tables will be updated in future ABSCO versions to maintain consistency with the latest HITRAN compilation. To date, the  $\text{CH}_4$  absorption coefficients have not been utilized within the L2 algorithm, but the bandpasses selected avoid the stronger methane lines. The ozone ( $\text{O}_3$ ) Chappuis-Wulf bands produce weak, quasi-continuum absorption within the  $\text{O}_2$  A band ( $2$  to  $4 \times 10^{-22} \text{ cm}^2/\text{molecule}$ ).  $\text{O}_3$  is not currently included in the OCO-2 forward model. The absorption by the other molecules shown in Figure 3-7, Figure 3-8, and Figure 3-9 is too small to produce measurable effects in the  $X_{\text{CO}_2}$  retrieval, and these molecules are currently omitted from the algorithm.

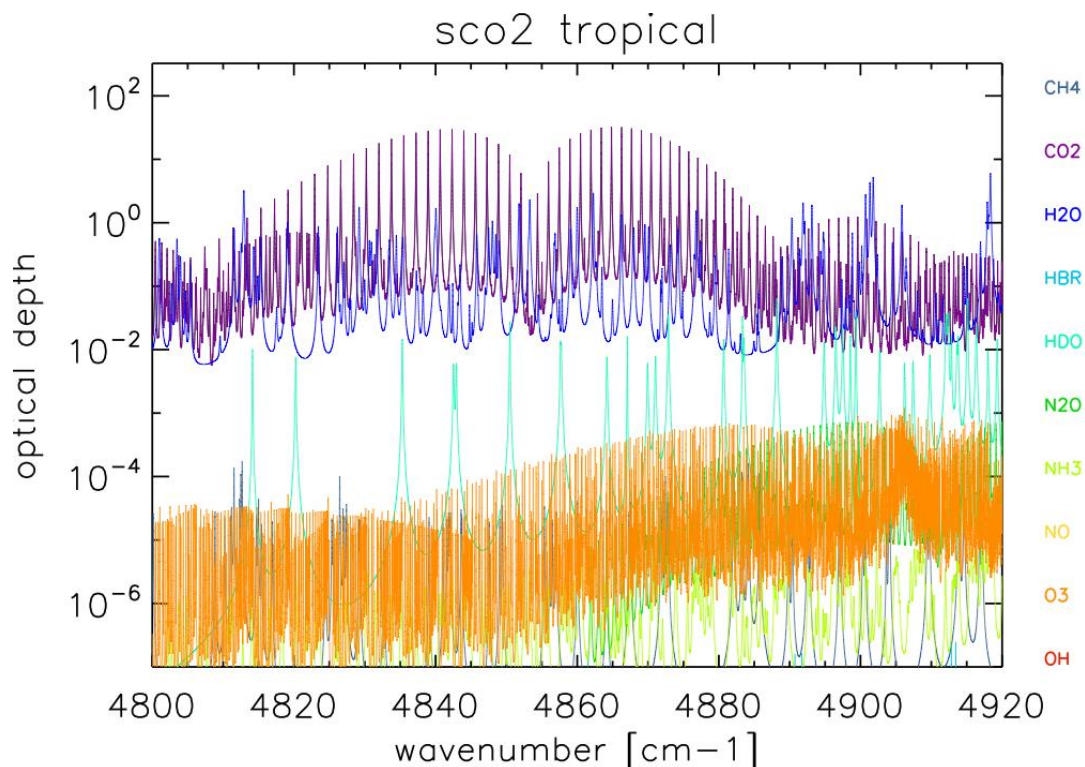




**Figure 3-7.** Optical depth contributions of different molecules in the O<sub>2</sub> A band for a US standard atmosphere.



**Figure 3-8.** Optical depth contributions for different gases in the 1.6 μm CO<sub>2</sub> band for a tropical standard atmosphere.



**Figure 3-9.** Optical depth contributions from different molecules in the 2.06  $\mu\text{m}$   $\text{CO}_2$  band for a tropical standard atmosphere.

### 3.2.1.3 Software Implementation

Due to the modular nature of the ABSCO calculation, different bands rely on different codebases to compute the absorption cross sections. In summary, the LABFIT codebase from Benner et al. [2016] provides  $\text{CO}_2$  cross sections. This implements databases from Benner et al. [2016] and Devi et al. [2016] for 2.0  $\mu\text{m}$  and 1.6  $\mu\text{m}$   $\text{CO}_2$ , respectively. The procedure and modeling assumptions of this approach are detailed at greater length in Thompson et al. [2012]. The same codebase has been extended to compute cross sections for  $\text{O}_2$  [Drouin et al., 2017; Payne et al., 2020]. We use the ABSCOI codebase, a routine adapted from the GFIT software package in use by TCCON [Wunch et al., 2011], for all  $\text{CH}_4$  and  $\text{H}_2\text{O}$  lines.

The table format is a four-dimensional (4D) array, allowing cross sections to be parameterized by temperature, pressure, and the volume mixing ratio of one other atmospheric gas ( $\text{H}_2\text{O}$ ). All tables are provided in the 4D format, although the calculations only incorporate broadening by  $\text{H}_2\text{O}$  of  $\text{O}_2$  and  $\text{CO}_2$ , and not for  $\text{CH}_4$ .

Table 3-2 below gives more detailed information on several key parameter sources.



**Table 3-2.** ABSCO v. 5.1 parameter sources.

Band	Spectral Range (cm <sup>-1</sup> )	Parameters	Source (ABSCO v5.1)
<b>0.76 μm</b>	12950–13190 cm <sup>-1</sup>	O <sub>2</sub> line shape, line mixing, CIA H <sub>2</sub> O-O <sub>2</sub> broadening H <sub>2</sub> O line parameters H <sub>2</sub> O continuum	Payne et al. [2020], Drouin et al. [2016] Drouin et al. [2014] HITRAN 2012 MT_CKD v3.2
<b>1.6 μm</b>	6120–6260 cm <sup>-1</sup>	CO <sub>2</sub> line shape, line mixing H <sub>2</sub> O-CO <sub>2</sub> broadening H <sub>2</sub> O line parameters H <sub>2</sub> O continuum	Devi et al. [2016] Sung et al. [2009] HITRAN 2012 MT_CKD v3.2
<b>2.06 μm</b>	4800–4890 cm <sup>-1</sup>	CO <sub>2</sub> line shape, line mixing Ad-hoc CO <sub>2</sub> absorption H <sub>2</sub> O-CO <sub>2</sub> broadening H <sub>2</sub> O line parameters H <sub>2</sub> O continuum	Benner et al. [2016] Thompson et al. [2012] Sung et al. [2009] HITRAN 2012 MT_CKD v3.2

### 3.2.1.4 Calculation of Gas Absorption Optical Depths in each Atmospheric Layer

Once we have the lookup tables of cross section on pressure  $P$ , temperature  $T$ , and wavelength  $\lambda$ , we must use the actual profiles of  $P$ ,  $T$ , and gas concentration to calculate optical depth for each layer and gaseous species.

Let us suppose we have a thick layer that goes from a lower pressure  $P_{bot}$  to an upper pressure  $P_{top}$  and we wish to calculate the optical depth of this layer for a certain species of gas. Let the dry-air volume-mixing ratio of this gas be  $q_{gas}$ , which is defined as the number of moles of the gas per mole of dry air. Further, let us call the absorption cross section of the gas (in  $m^2 mol^{-1}$ )  $k(P, T)$  at a particular wavenumber. Because the temperature  $T$  is an intrinsic function of the pressure  $P$  for the given profile, we write the gas absorption coefficient for this specific layer as  $k(P)$ .

In an infinitesimal layer, it can be shown that the number of molecules of *dry air* (against which the volume mixing ratio of any gas is defined) can be written:

$$N_{dry} dz = -\frac{dP}{gM_{dry}} (1 - q) \quad (3-13)$$

where  $dZ$  is the layer width in meters,  $N_{dry}$  is the number of moles of dry air per cubic meter,  $g$  is the local gravitation acceleration,  $q$  is the local specific humidity, and  $M_{dry}$  is the molar mass of dry air in  $kg mol^{-1}$ . Thus, the total optical depth of a thick layer can be written:

$$\tau_{gas} = \int_{P_{top}}^{P_{bot}} q_{gas} \frac{k(P)[1 - q(P)]}{g(P)M_{dry}} dP \quad (3-14)$$

Note that we have expressly allowed  $g$ ,  $q$ ,  $q_{gas}$ , and  $k$  to vary as a function of pressure across this layer. Unfortunately, we only know each quantity at the boundary of the layer in question. Therefore, we must make assumptions about how each varies across the layer in order to evaluate the integral. Initially, each quantity was simply evaluated at the center of the layer assuming a linear variation; however, this approach can cause small biases in the total optical depth of a given gas and can lead to corresponding biases in retrieved  $X_{CO_2}$ . This is mostly due to

layers at the top of the atmosphere that span close to two orders of magnitude in pressure (say from 1 to 100 mbar)—the variation across this layer in  $k$  is so great (and quite nonlinear) that we cannot simply ignore it.

A straightforward approach that largely eliminates this bias is to divide each layer into a number of sublayers. Then the optical depth for each sublayer is evaluated according to the approximation:

$$\tau_{\text{sub}} = \frac{q_{\text{gas}, \text{cen}} (1 - q_{\text{cen}})}{g_{\text{cen}} M_{\text{dry}}} \cdot K(P_{\text{cen}}, T_{\text{cen}}) \cdot \Delta P \quad (3-15)$$

where the *cen* subscript indicates a quantity to be evaluated at the center of the sublayer; all such quantities are evaluated using linear interpolation in pressure.

For each sublayer,  $k(P_{\text{cen}}, T_{\text{cen}})$  is evaluated using 2D linear interpolation in both pressure and temperature. The optical depth of the full layer is then simply the sum of the optical depths of its component sublayers.

The L2 FP code subdivides each model layer into 10 sublayers of equal pressure width, plus adding additional sublayers at each of the meteorology pressure levels. The 10 sublayers capture the variation of the ABSCO table with pressure, while using the meteorology pressure levels captures the full temperature profile (which is specified on the meteorology pressure levels). Once the sublayer optical depths are evaluated, they can be summed according to Simpson's rule to obtain the total gas optical depth of each layer (and for each species of gas). This algorithm for integration was compared to a full (and much slower) integration using the Gnu Scientific Library, giving a maximum difference of 0.6%.

Currently, the L2 code makes this calculation for O<sub>2</sub> in the O<sub>2</sub> A band and for both CO<sub>2</sub> and H<sub>2</sub>O in the weak and strong CO<sub>2</sub> bands. Abundances of the minor isotopologues of each are a fixed fraction of the total dry-air volume-mixing ratio, set inside the spectroscopic lookup tables.

### 3.2.1.5 Rayleigh Scattering Cross Sections

Rayleigh scattering is accounted for in all OCO bands, though it is most important in the O<sub>2</sub> A band. The Rayleigh scattering cross section is calculated using the following classic model:

$$\sigma_R = \frac{24\pi^3 (n_s^2 - 1)^2}{N_s^2 \lambda^4 (n_s^2 + 2)^2} \frac{6 + 3\rho}{6 - 7\rho} \quad (3-16)$$

where  $\lambda$  is the wavelength,  $\rho$  is the depolarization factor,  $N_s$  is the number density of air at some pressure and temperature, and  $n_s$  is the index of refraction of standard air *at the same pressure and temperature*, and  $\rho$  is the depolarization factor, which describes the effect of molecular anisotropy. According to Lorentz-Lorenz theory, the product of the two terms involving these two variables is independent of pressure and temperature. We take  $N_s = 2.687 \cdot 10^{19}$  molecules/cm<sup>3</sup> at 273.15 K and 1013.25 mbar. The factor  $24\pi^3/N_s^2$  may be calculated to be  $1.031 \cdot 10^{-20}$  m<sup>2</sup> μm<sup>4</sup> per molecule, such that specifying  $\lambda$  in μm will result in the cross section having units of m<sup>2</sup>/molecule.

The index of refraction for air at standard pressure and 273.15 K is taken from Allen [1964], using the following parameterization:

$$n_s = 1 + a * (1 + b\lambda^2) \quad (3-17)$$

where  $\lambda$  is in microns,  $a = 2.871 \times 10^{-4}$ , and  $b = 5.67 \times 10^{-3}$ . The depolarization factor is taken to be a constant over all the OCO bands, with a value of 0.0279 as suggested by Young [1980]. The Rayleigh phase matrix including polarization is a standard formulation [e.g., Hansen and Travis, 1974].

In order to calculate the Rayleigh optical depth  $\tau_R$  for a given layer, it is necessary to multiply the cross section by the total number of air molecules per  $m^2$  in the layer. For a given layer with pressure width  $\Delta P$  given in Pa and mean gravity  $g$  in  $m/s^2$ , this is given approximately by:

$$N_{tot} \Delta z = \frac{N_A \Delta P}{g M_{dry}} \left( 1 + \frac{1 - \epsilon}{\epsilon} q \right) \quad (3-18)$$

where  $N_A = 6.02214 \times 10^{23}$  is Avogadro's number,  $M_{dry} = 28.96 \text{e-}3$  is the molar mass of dry air in kg/mole, and  $\epsilon$  is the ratio of the molar mass of dry air to that of water vapor ( $\epsilon \approx 0.622$ ). Note that this represents the total number of air molecules in a layer, and therefore includes both the wet and dry contributions.

### 3.2.1.6 Cloud and Aerosol Optical Properties

The single-scattering optical properties (cross sections, single-scattering albedos, and scattering phase matrices) of cloud and aerosol particles are precomputed and tabulated. Currently, the OCO-2 retrieval algorithm includes five different types of scattering particles for each sounding. The first two are liquid water and ice clouds, described in the next paragraph. Another two are the two dominant aerosol types, set a priori to live in the lower troposphere (see section 3.3.2.3 for details), selected from the five available in the MERRA model. The last type is a sulfate aerosol that is confined to live in the upper atmosphere (section 3.3.2.3), and uses the same optical properties as the tropospheric sulfate aerosol type.

The wavelength-dependent optical properties for liquid water clouds were calculated from Mie theory assuming a Gamma particle size distribution and an effective radius of 8  $\mu m$ . The optical properties of ice clouds with a single effective diameter of 70  $\mu m$  were obtained from a database of ice crystal optical properties, compiled for Moderate-Resolution Imaging Spectroradiometer (MODIS) Collection-6 [Baum et al., 2014]. This database makes use of severely roughened ice particles and has more realistic phase functions than that from Collection-5, which was used in previous ACOS retrieval versions. This database also has the advantage that it was calculated at high wavelength resolution, and therefore did not have to interpolate across important spectral features, as occurred in the previous version. Figure 3-10 shows the six independent phase matrix elements of the 70  $\mu m$  severely roughened ice crystals at a wavelength of 765 nm. To save on processing time, we employ delta-M truncation [Wiscombe, 1977] with 5000 moments. As shown in Figure 3-10, there is no visible difference between the exact phase function and the truncated one.

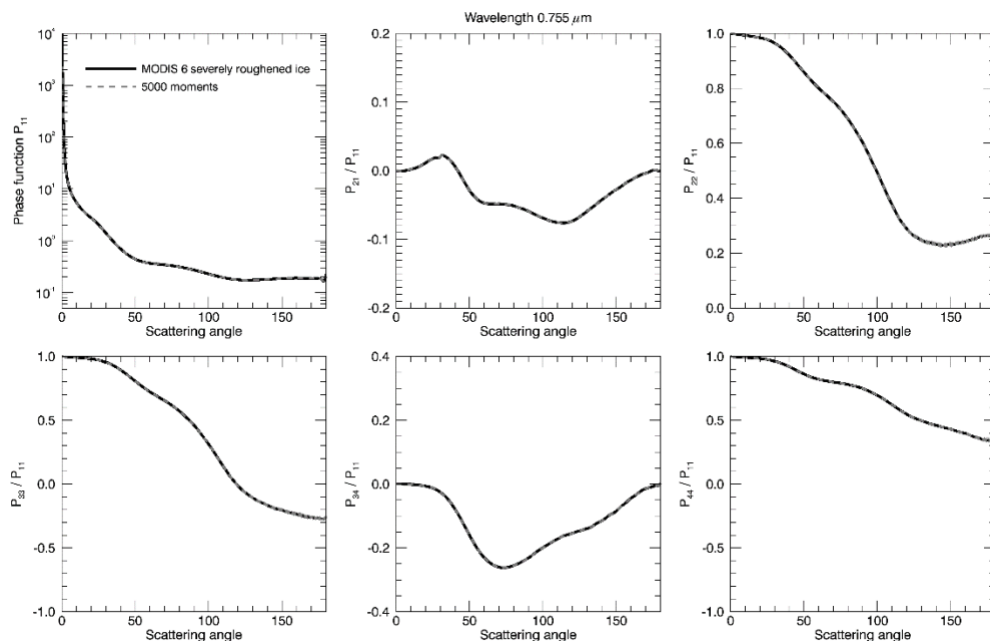
The five aerosol types are based on the aerosol types used in the Modern-Era Retrospective Analysis for Research and Applications (MERRA) reanalysis system: dust, smoke, sea salt, sulfate aerosol, organic carbon, and black carbon. In that system, the sea salt, carbon types, and smoke are all hygroscopic (dependent upon relative humidity). Further, five differently-sized bins each of dust and sea salt are separately tracked in the model. These bins plus the hygroscopicity lead to a huge range of effective aerosol types and associated optical properties.

In the L2 retrieval, we wish to use the types most commonly associated with a given location and month, but representing that huge level of detail is likely unnecessary. Therefore, we find mean relative humidity values for each subtype and calculate optical properties only for those relative humidity values. Second, we composite the different dust and sea salt bins into a single dust type and a single sea salt type. The L2 algorithm has flexibility of the number of aerosol types used. The default configuration uses the two most prevalent types, with the optical depth of these two types scaled to match the optical depth of all five types in MERRA. Below, we first describe how we choose the mean relative humidity values, and then we give the mathematics that describe the compositing of the aerosol optical properties.

### 3.2.1.6.1 Aggregation of MERRA Aerosol Subtypes

In choosing how to combine different subtypes into an average aerosol type, we use the fact that the densities of some aerosol types or subtypes are highly correlated with one another, such that we can form a mixture of the two. The only things that remain are (1) to determine the relative fractions of each type to use (in terms of fraction of the aerosol density), and (2) to determine “typical” relative humidity (RH) values to use, as some of the types have optical properties that depend on RH.

In the following analysis, we selected 100 3-hour MERRA reanalysis files at random. This was done a couple of times to make sure the results were fairly stable. In Table 3-3, we give the results for the density weightings that were chosen for L2 algorithm, as well as chosen typical



**Figure 3-10.** Independent phase matrix elements for the cloud ice particles used in the L2 retrieval algorithm. Shown are both the exact phase functions, as well as the M=5000 delta-M.

values of RH to use. The latter was based primarily on weighting by optical depth in the O<sub>2</sub> A band.

**Table 3-3.** Density weightings and RH values chosen for the L2 algorithm.

Aggregate Name	Subtypes	Density Fractions [%]	RH values [%]
Dust	DU Bins 1, 2, 3, 4, 5	15, 40.5, 30.5, 11.5, 2.5	N/A
Sea Salt	SS Bins 1, 2, 3, 4, 5	0.2, 2.2, 19.4, 60.6, 17.6	88 (all)
Black Carbon	BC Phobic, Philic	21, 79	54
Organic Carbon	OC Phobic, Philic	16, 84	77
Sulfate Aerosol	SU (Type 1)	100	70

### 3.2.1.6.2 Composite Optical Properties of the Aggregate MERRA Aerosols

The MERRA aerosol optical properties are a function of relative humidity and effective radius for a number of types, where each type has potentially a number of size bins that are kept separate. Two or more of these subtypes can be put together as a composite type. We are given the complete optical properties for the individual aerosol subtypes by colleagues at NASA GSFC. Let us first define the following symbols:

- $Q$ : extinction efficiency
- $S$ : physical cross section per unit volume in  $\text{m}^{-1}$
- $\beta$ : volume extinction coefficient (cross section per unit volume) in  $\text{m}^{-1}$
- $k$ : mass extinction coefficient in  $\text{m}^2/\text{kg}$
- $\lambda$ : wavelength
- $\omega$ : single scattering albedo
- $f$ : fraction of density
- $\rho$ : mass of aerosol per  $\text{m}^3$  of air
- $g$ : asymmetry parameter

For any aerosol type or mixture, we have:

$$\beta(\lambda) = k(\lambda)\rho \quad (3-19)$$

where  $\beta$  and  $k$  can be for extinction ( $e$ ) or scattering ( $s$ ). Let our composite aerosol type have  $\beta(\lambda)$ ,  $k(\lambda)$ , and  $\rho$ . Let us define the fractionation between subtypes in terms of aerosol density. We do this as follows for two contributing subtypes (1 and 2), but this formalism is trivially expanded to more subtypes:

$$\rho = \rho_1 + \rho_2 = f_1\rho + f_2\rho \quad (3-20)$$

and  $f_1 + f_2 = 1$ . The total volume extinction coefficient  $\beta$  is just the sum of the individual  $\beta$ s:

$$\begin{aligned} \beta_e &= \beta_{e,1} + \beta_{e,2} \\ &= k_{e,1}\rho_1 + k_{e,2}\rho_2 \\ &= k_{e,1}f_1\rho + k_{e,2}f_2\rho \\ &= (k_{e,1}f_1 + k_{e,2}f_2)\rho \end{aligned} \quad (3-21)$$

From Equation (3-19) and Equation (3-21), we see that the composite mass extinction coefficient is given by:

$$k_e = k_{e,1}f_1 + k_{e,2}f_2 \quad (3-22)$$

So, this implies that we can easily calculate mass extinction coefficient  $k_e$  for the composite type. The mass scattering coefficient  $k_s$  is calculated similarly. Finally, the composite single scattering albedo  $\omega$  is just the ratio of the composite scattering coefficient to the composite extinction coefficient:

$$\omega = \frac{k_s}{k_e} = \frac{k_{s,1}f_1 + k_{s,2}f_2}{k_{e,1}f_1 + k_{e,2}f_2} = \frac{k_{s,1}f_1 + k_{s,2}f_2}{k_e} \quad (3-23)$$

Finally, we want to determine phase function quantities, such as  $g$ , and phase matrix expansion coefficients. All phase function quantities will composite exactly like  $g$ . Therefore, it is sufficient to show how to find the composite  $g$ , and this can be applied to all phase function objects (expansion coefficients, etc.). It is straightforward to show that:

$$g = \frac{k_{s,1}g_1f_1 + k_{s,2}g_2f_2}{k_{s,1}f_1 + k_{s,2}f_2} = \frac{k_{s,1}g_1f_1 + k_{s,2}g_2f_2}{k_s} \quad (3-24)$$

Note that we have not calculated any  $Q$ -type quantities yet (extinction efficiency, etc.). Formally, these are never needed in the L2 FP algorithm, as the wavelength-dependence of  $Q$  ( $e$  or  $s$ ) is exactly the same as for  $k$  ( $e$  or  $s$ ) for any given type or subtype. However, they are the quantities we typically specify in our optical properties files, and so we discuss them here.

Physically,  $Q$  is the extinction cross section per unit volume divided by the physical cross section per unit volume:

$$Q = \frac{\beta}{s} \quad (3-25)$$

The effective  $Q$  over a full size distribution of spherical particles is given by:

$$Q = \frac{4}{3}k \cdot R_{\text{eff}} \cdot \rho_{\text{bulk}} \quad (3-26)$$

where  $\rho_{\text{bulk}}$  is the bulk density of the aerosol material, and in general may be a function of the relative humidity (but not the wavelength!). GSFC gives  $Q$ ,  $\beta$ , and  $R_{\text{eff}}$  for each aerosol type and as a function of relative humidity. Therefore, the bulk density they use is easily calculated for each of their five main types as a function of RH. For dust, they use a bulk density of 2650 kg/m<sup>3</sup>, except for the first size bin where they seem to have used a bulk density of 2500 kg/m<sup>3</sup>. Using the equation for  $Q$  and knowing that  $S$  of a composite type is simply the sum of the  $S$ s of the contributing subtypes, it is easily shown that the composite  $Q_e$  can be calculated from the individual  $Q_e$ s according to:

$$Q_e = \frac{k_e}{\frac{k_{e,1}f_1}{Q_{e,1}} + \frac{k_{e,2}f_2}{Q_{e,2}}} \quad (3-27)$$

and similarly for  $Q_s$ .

The extinction efficiency, single scattering albedo, and asymmetry parameters of the five aerosol types and two cloud types are shown in Figure 3-11. The asymmetry parameter represents the degree of forward scattering in the intensity phase function. All scattering types include a fully polarized treatment of the scattering phase function using a standard formalism [de Rooij and van der Stap, 1984]. The scattering phase matrix for ice was shown in Figure 3 11.

### 3.2.2 Solar Spectrum Model

The top-of-atmosphere solar flux spectrum used in the OCO-2 FP forward model is generated by convolving a high-resolution pseudo-transmittance spectrum with an empirical, radiometrically-calibrated solar continuum spectrum. Doppler shifts due to variations in Sun-Earth distance and Earth rotation are individually corrected for each sounding.

#### 3.2.2.1 Pseudo-Transmittance Spectrum

The high-resolution pseudo-transmittance spectrum is derived by combining an empirical solar line list with a simple, analytic solar line shape function, both of which were developed by G. Toon and colleagues [Toon et al., 1999]. This approach offers advantages over solar spectra generated by other methods in this application. For example, since the line list and solar line shape are predefined, the pseudo-transmittance spectrum can be pre-calculated at high resolution and stored for quick table lookup during the retrievals. By using a line list and line shape, data from several sources can be combined to increase the spectral coverage and remove the instrumental and atmospheric effects and provide a selectable-resolution, exo-atmospheric solar pseudo-transmittance spectrum. Additionally, since both instrumental and atmospheric effects are removed during the line list fitting process, they can be ignored in the OCO-2 code.

The line list contains more than 35,000 lines and covers the 600–25,000  $\text{cm}^{-1}$  region. For the near-infrared (NIR) region covered by OCO-2, the line list is based primarily on ground-based Kitt Peak spectra, with the exception of the  $\text{O}_2$  A-band regions, which are based primarily on Denver University balloon spectra. Additionally, some ground-based TCCON spectra from Park Falls were used. The spectrum has been validated against Izana TCCON spectra in the NIR as well as GOSAT spectra at lower resolution.

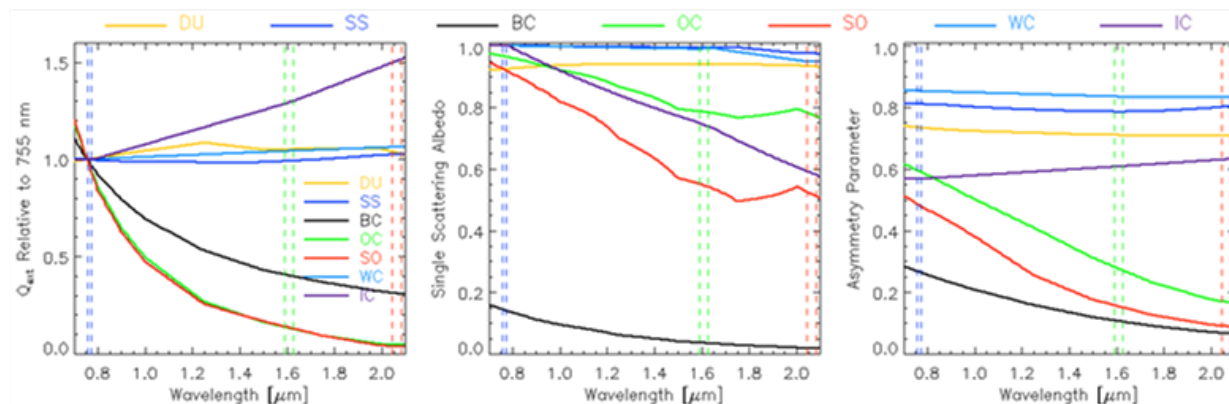
The empirical solar line shape function uses a Doppler shape near the center with exponentially decaying wings. The line shape does not have a physical basis but has been found to provide a reasonable approximation of the observed solar line shapes. Readers should contact G. Toon directly for more detailed information regarding the derivation and implementation of the line list and line shape function.

#### 3.2.2.2 Solar Continuum Model

The solar continuum model was updated in the version 10 data product. This continuum was originally derived by fitting a 9th order polynomial across the OCO-2 spectral range (0.757 to 2.08  $\mu\text{m}$ ) to a low-resolution extra-terrestrial solar spectrum. Earlier OCO-2 products used a continuum fit to the solar spectrum acquired by the ATmospheric Laboratory for Applications and Science 3 (ATLAS 3) Solar Spectrum (SOLSPEC) instrument, which flew on Space Shuttle Atlantis in 1994 [Thuillier et al., 2003]. However, recent measurements from the SOLar SPECTrometer (SOLSPEC) instrument in the SOLAR payload [Meftah et al., 2017] and the Total

and Spectral Solar Irradiance Sensor (TSIS) Spectral Irradiance Monitor (SIM) instruments on the International Space Station (ISS) showed significant discrepancies with the ATLAS 3 SOLSPEC estimates.

For the version 10 product, the TSIS SIM results (E. Richard, personal communication) were adopted as the new standard. To update the continuum in each OCO-2 spectral channel, the existing continuum values were first scaled by a multiplicative offset and slope. The modified continuum was then convolved with the high-resolution pseudo-transmission spectrum to yield a high-resolution solar spectrum. This high-resolution spectrum was then convolved with TSIS-IM spectral response function and the resulting (low-resolution) spectrum was compared to the TSIS-SIM observations. This process was repeated until the modeled and observed solar flux spectra matched to within  $\sim 0.5\%$  throughout each of the OCO-2 spectral ranges. The resulting spectra had continuum values reduced by  $\sim 1.3\%$  near  $0.760\mu\text{m}$ ,  $3\%$  near  $1.610\mu\text{m}$ , and  $6.5\%$  near  $2.060\mu\text{m}$ . These changes produced no significant global bias in the retrieved  $X_{\text{CO}_2}$  estimates. They did change the retrieved surface pressures by  $-0.2\text{ hPa}$  over land and  $+0.2\text{ hPa}$  over ocean. They also reduced the retrieved surface wind speed estimates for ocean glint observations by  $\sim 1\text{ m/sec}$  because these wind speeds are derived from a Cox-Munk model constrained by the  $2.06\mu\text{m}$  radiances, which were altered the most by this update.



**Figure 3-11.** Aerosol and cloud optical properties assumed in the L2 code. **Left:** Extinction Efficiency relative to that at 755 nm. **Middle:** Single Scattering Albedo. **Right:** Asymmetry Parameter. DU: Dust, SS: Sea Salt, BC: Black Carbon, OC: Organic Carbon, SO: Sulfate, WC: Water Cloud, IC: Ice Cloud. The three OCO-2 bands are demarcated by the dashed vertical lines.

For use in the OCO-2 algorithm, the modified solar continuum in each channel is interpolated to the spectral grid of the high-resolution pseudo-transmittance spectrum, and convolved with that spectrum. We assume that the solar continuum is invariant over time. However, the baseline continuum is specified for a Sun-Earth distance of 1 Astronomical Unit (AU;  $1\text{ AU} = 149,597,871\text{ km}$ ). It is therefore necessary to scale the spectrum with the square of the seasonally-varying Earth-Sun distance, which has been taken from <http://eclipse.gsfc.nasa.gov>.

### 3.2.2.3 Solar Doppler stretch

The relative velocity of the Earth and the Sun caused by the Earth's rotation and the ellipticity of the Earth's orbit about the Sun produces daily and seasonal Doppler shifts in the solar lines with respect to the telluric lines. The Doppler shift  $\Delta\lambda$  for a wavelength  $\lambda$  is calculated assuming a small (non-relativistic) relative velocity  $v_{\text{rel}}$ :



$$\frac{\Delta\lambda}{\lambda} = \frac{-v_{rel}}{c} \quad (3-28)$$

where  $c$  is the speed of light.

The relative velocity is given by the sum of the Earth-Sun radial velocity and the Earth rotational velocity component. The Earth-Sun velocity is calculated from the derivative of a 6th order polynomial fit of solar distance derived from NASA Reference Publication 1349—Twelve Year Planetary Ephemeris: 1995 – 2006. Note that this correction does not include the Doppler shift due to the spacecraft’s  $\sim 7$  km/sec velocity around the Earth, which affects solar and telluric lines equally. That correction is made within the instrument model.

### 3.2.3 Radiative Transfer

The radiative transfer (RT) module encapsulates all the physics associated with the modification of the solar radiation during its passage through the atmosphere and reflection by the surface. This is greatly complicated by the fact that some of the radiation is reflected back to space (by cloud/aerosol) before hitting the ground, causing the average photon path to be wavelength-dependent, as well as by the fact that reflected radiation is strongly polarized, especially in glint mode over ocean.

Because OCO-2 is polarization-sensitive and the radiation can be highly polarized, it is essential that the forward model calculations take this into account and calculate the fully-polarized radiation. We describe the radiance at a given wavelength incident upon the instrument by a three-element Stokes vector  $\{I, Q, U\}$ . We neglect the circular polarization component because it is generally negligible and OCO-2 is not sensitive to it. Traditionally, a fully-polarized RT code such as vector LIDORT (or VLIDORT) would be used for the calculation, but this is too computationally expensive. Instead, we resort to a piecemeal approach. First, we approximate the Stokes vector as:

$$I = I_1 + I_{ms} + I_{corr} \quad (3-29)$$

$$Q = Q_1 + Q_2 \quad (3-30)$$

$$U = U_1 + U_2 \quad (3-31)$$

where the subscript  $1$  denotes first order of scattering (and includes the direct beam reflected from the surface),  $2$  denotes second order of scattering,  $ms$  denotes all orders of scattering above the first, and  $I_{corr}$  represents a polarization correction to  $I_{ms}$ . The various terms are then calculated as follows:

1.  $I_1$ ,  $Q_1$ , and  $U_1$  are calculated by a fast, fully-polarized single order of scattering module
2.  $I_{ms}$  is calculated by LIDORT, a multiple-scattering scalar RT code
3.  $Q_2$ ,  $U_2$ , and  $I_{corr}$  are calculated by a fully-polarized, second order of scattering technique called 2OS.

Our goal is a spectrum of  $I$ ,  $Q$ , and  $U$  at each of the  $\sim 40,000$  high-resolution spectral points in order to convolve with the instrument response. However, even with the savings of not using a vectorized, all-orders-of-scattering code, a direct brute-force process is still too expensive. Therefore, a method dubbed the “Low Streams Interpolator” is employed to drastically speed up the RT process and avoid the direct, brute-force approach.

### 3.2.3.1 LIDORT

The FP algorithm uses LIDORT [Spurr et al., 2001; Spurr, 2002] to solve the RT equation (RTE). LIDORT is a linearized discrete ordinate RT model that generates radiances and Jacobians (derivatives of the radiance with respect to atmospheric and surface properties) simultaneously. The Jacobians are computed by an internal perturbation analysis of the complete discrete ordinate solution to the RTE. LIDORT is a quasi-spherical model; the direct beam and line of sight attenuations are treated for a curved atmosphere while the multiple scattering is considered to be locally plane parallel.

### 3.2.3.2 Polarization Correction: The Two Orders of Scattering Model

The OCO-2 FP code uses LIDORT as the scalar RT model. However, to fully describe the state of light observed by the detector, polarization has to be considered. Calculations of multiple scattering with full treatment of polarization are computationally very expensive. Since multiple scattering is depolarizing, it is reasonable to expect that the polarization could be accounted for by a low-order scattering approximation. Natraj and Spurr [2007] extended the scalar Kawabata and Ueno [1988] model to compute the first two orders of scattering (2OS) for vertically inhomogeneous scattering media with polarization included. To enable accurate computations for the range of solar viewing angles encountered by OCO-2, atmospheric transmittances for the incoming solar beam are treated for a curved spherical-shell atmosphere [Spurr, 2002]. For the glint and nadir modes of operation, there is also a correction for the sphericity along the line of sight. Polarization induces a change in the intensity; to account for this, we compute a correction to the scalar intensity. The 2OS model simultaneously computes both the simulated backscatter radiance (and intensity correction) and any number of associated weighting functions (partial derivatives of the radiance with respect to retrieved and other atmospheric and surface properties). The 2OS computation is an order of magnitude faster than a full multiple-scattering scalar calculation and two orders of magnitude faster than a vectorized multiple-scattering computation.

#### 3.2.3.2.1 Theory

The formulation of the 2OS model is based on the invariant imbedding theory of RT. The contribution from the  $p^{th}$  order of scattering is related to that from the  $p-1^{th}$  order of scattering by an integro-differential equation. Starting from direct transmission, the expressions for subsequent orders of scattering can be derived. This is particularly straightforward for the first two orders of scattering. A Fourier series expansion is done to solve the problem efficiently. The expressions for the Stokes parameters  $I$ ,  $Q$ , and  $U$  are as follows:

$$I_{OCO} = I_{sc} + I_{corr} \quad (3-32)$$

$$Q_{OCO} = Q_1 + Q_2 \quad (3-33)$$

$$U_{OCO} = U_1 + U_2 \quad (3-34)$$

where the subscripts  $OCO$ ,  $sc$ ,  $I$ , and  $2$  are, respectively, the OCO-2 measurement, scalar computation (from LIDORT), and first and second order computations (from the 2OS model).  $I_{corr}$  is the intensity correction computed by the 2OS model. Equation (3-32) implies that the LIDORT is employed in multiple-scattering mode (single scattering is computed using the 2OS model). The reasons for this will become apparent in Section 3.2.3.3.

### **3.2.3.2.2 Evaluation**

The OCO-2 FP RT model has been tested against the full multiple scattering vector model VLIDORT [Spurr, 2006] for the OCO-2 nadir, glint, and target modes of operation.

Computations were performed for different solar zenith angles, surface types, and aerosol loadings [Natraj et al., 2008]. The radiance errors using the hybrid LIDORT-2OS model are an order of magnitude (or more) smaller than the errors arising from the use of LIDORT alone. In addition, a linear error analysis study showed that the errors in the retrieved  $X_{CO_2}$  using the hybrid model are much lower than the measurement noise and smoothing errors. On the other hand, the scalar model alone induces  $X_{CO_2}$  errors that could dominate the retrieval error budget.

### **3.2.3.3 Acceleration of the RT Calculations**

#### **3.2.3.3.1 Low-Streams Interpolation**

The scalar RT code used to calculate the multiple-scattering component of the total intensity ( $I_{ms}$  from Equation (3-29)), LIDORT, relies on a quadrature-based approach wherein the code effectively only treats radiation coming from specific discrete zenith angles or “streams,” and interpolates between these angles to find the radiation at an arbitrary angle. In standard nomenclature, a code is said to be run with  $N$  streams total, where  $N/2$  is the number of streams in either the upper or lower hemisphere. The computational burden in theory goes as  $N^3$ , because the codes largely rely on multiplication and inversion of  $N \times N$  matrices. However, often there is other significant computational overhead, which lessens the extreme dependence on  $N$  to be more like  $N^2$ , especially for lower numbers of streams.

The number of streams required to achieve a given accuracy depends on the type of atmosphere in question. To obtain sufficient accuracy for most scenes and viewing geometries, eight to 32 streams are required. However, doing this in a line-by-line approach is cost-prohibitive. To speed this process, we employ the method of low-streams interpolation (LSI), fully explained in O’Dell [2010]. In a nutshell, this process allows one to perform just a handful (five to 20 per band) of high-accuracy calculations, alongside a low-accuracy RT calculation for every monochromatic wavelength. The monochromatic high-accuracy calculations can then be reconstructed with very low errors using the LSI method. In addition, because the technique is so simple (involving just a couple of very simple equations), it was very straightforward to linearize such that it can handle Jacobians as well. In our particular setup, we have:

- Low accuracy (monochromatic) calculations: two-stream scalar RT plus polarized single-scattering code
- High accuracy calculations: 16-stream scalar RT plus polarized 2OS calculations (Section 3.2.3.2)

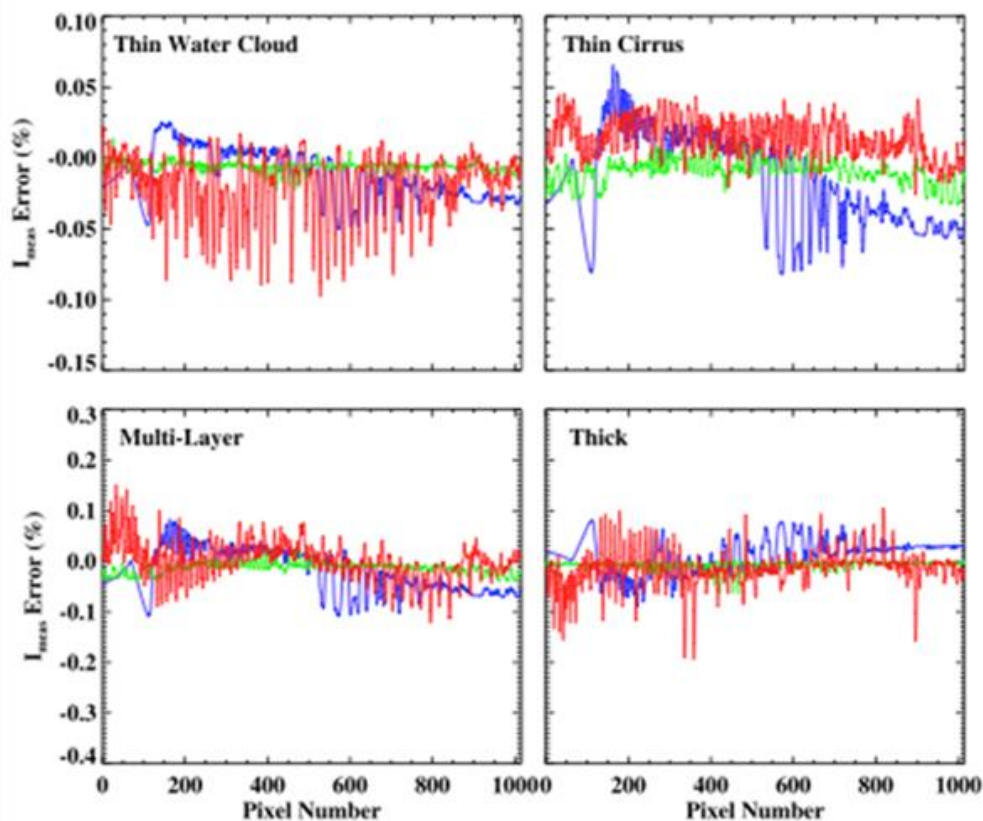
The errors from a similar setup are shown in Figure 3-12 (taken from O’Dell [2010]). Internal tests have shown that the errors in  $X_{CO_2}$  from this approach tend to be random and typically less than 0.3 ppm.

### 3.2.3.3.2 Non-Uniform Spectral Sampling

Although the errors in the LSI are acceptable, the resulting RT is still slower than necessary. The low-accuracy monochromatic RT calculations typically dominate the overall calculation time. Therefore, we found it acceptable to “skip” some of these wavelengths and simply interpolate from neighboring spectral points to obtain these values. Recall that we perform our RT calculations on a  $0.01\text{ cm}^{-1}$  spaced grid for all three NIR bands. It turns out this this spacing most over-samples the  $\text{O}_2\text{ A}$  band, though there is some oversampling in each band.

We use a simple approach to determine which spectral points can be skipped. It is a semi-automated process that requires monochromatic radiance errors from interpolation to be less than some threshold, typically 0.1%. We have investigated different interpolation techniques and generally find that linear interpolation is comparable in accuracy to higher order interpolation schemes, as shown in Figure 3-12 below.

Implementing “non-uniform sampling” leads to speed savings of about 50% in the low-accuracy, monochromatic RT calculations. The savings is highest in Band 1 and is almost negligible in Band 3, which is the least oversampled. The L2 retrieval realizes an overall speed savings of ~20%, though this depends on the details of the low- and high-accuracy RT



**Figure 3-12.** Percent errors in OCO-simulated TOA radiances due to the LSI method for four different cases. Each case includes one or multiple clouds. The three OCO bands are shown in blue ( $\text{O}_2\text{ A}$ ), green (weak  $\text{CO}_2$ ), and red (strong  $\text{CO}_2$ ). Taken from O'Dell [2010], Figure 8.

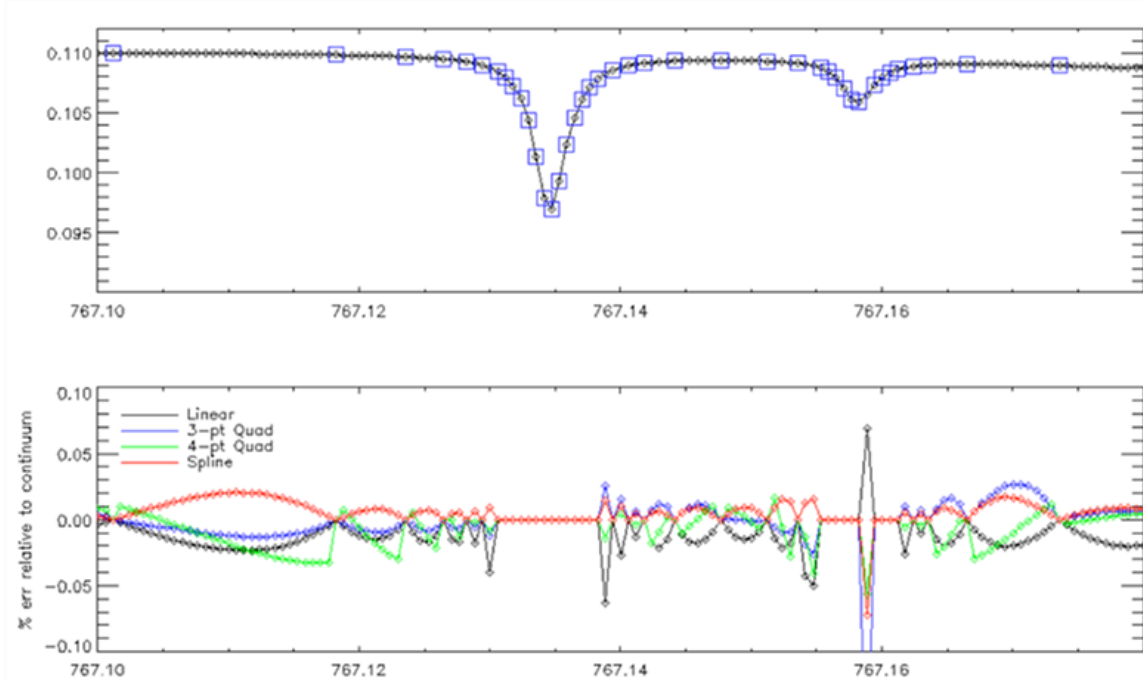
calculation setup. The errors in retrieved  $X_{CO_2}$  due to non-uniform spectral are virtually always less than 0.3 ppm, and are typically less than 0.1 ppm.

#### 3.2.3.4 Fluorescence Radiative Transfer

The fluorescence state vector elements are discussed in Section 3.3.2.4.3. This shows to how to calculate the SIF signal at the surface, which is taken to be isotropic and unpolarized. As discussed in that section, the emitted radiance due to fluorescence is modeled in the  $O_2$  A band as linear in wavenumber. This would break down over a large spectral range, but is appropriate for the narrow  $O_2$  A band from 759 – 772 nm. To propagate the SIF signal at the surface,  $SIF_{gnd}$ , to the top of atmosphere (TOA), we only take into account absorption by molecular oxygen:

$$SIF_{TOA} = SIF_{gnd} \cdot \exp\left(\frac{-\tau_{oxygen}}{\mu}\right) \quad (3-35)$$

where  $\tau_{oxygen}$  is the total zenith optical depth due to  $O_2$ , and  $\mu$  is the cosine of the sensor zenith angle; the equation is implicitly a function of (monochromatic) wavelength.  $SIF_{TOA}$  is then simply added to the upwelling solar radiance at TOA. For simplicity, we have ignored scattering and extinction due to clouds and aerosols, as we typically only retrieve  $X_{CO_2}$  in reasonably clear scenes. In reality, the extinction sink and scattering source of radiance due to clouds and aerosols often partially cancel, so this approximation is typically adequate for our purposes.



**Figure 3-13.** Illustration of non-uniform sampling approach. **Top:** Training. From a full monochromatic spectrum (black diamonds) certain spectral points are selected for RT calculations (blue squares). The remaining points are calculated from linear interpolation of the selected points. **Bottom:** Interpolation errors in the non-selected points for different interpolation schemes. Linear interpolation is used because its errors are comparable to higher-order interpolation schemes.

### 3.2.4 Instrument Model

The RT model produces high-resolution ( $\sim 0.01 \text{ cm}^{-1}$ ) synthetic spectra of  $I$ ,  $Q$ , and  $U$  for a specified observing geometry and surface/atmospheric state. Before these synthetic spectra can be compared to the calibrated, geolocated spectral radiances, the synthetic  $I$ ,  $Q$ , and  $U$  must then be corrected for instrument polarization, corrected for Doppler shift, and convolved with the instrument line shape (ILS) function. Each of these steps is defined below. The instrument model performs these steps.

Pixel-wavelength mapping: the wavelength  $\lambda$  for detector pixel  $i$  and spectrometer  $k$  is given by:

$$\lambda_i = \sum_{n=1}^N d_k^n \times i^{n-1} \quad (3-36)$$

where  $d$  are the  $N$  dispersion coefficients for spectrometer. For OCO-2,  $i$  ranges from 1 to 1016 for each spectral band and there are six coefficients for each band. These relationships also vary for each of the eight footprints within a band. The dispersion coefficients are stored in each Level-1B file. In the V8 product, the retrospective L1B dispersion is modified to include corrected coefficients for the first two terms ( $d^0$  &  $d^1$ ). These corrections are derived from the L2 retrieval algorithm and account for Doppler and thermo-mechanical shifts observed, in flight, in the dispersion (refer to the Level 1B Algorithm Theoretical Basis document).

#### 3.2.4.1 Instrument Line Shape Function

The ILS function describes the response of an individual OCO-2 detector pixel to light with a wavelength  $\lambda$ . Mathematically, this can be described as the product of the monochromatic intensity  $I$  with the ILS:

$$I(\lambda) = \int_{-\infty}^{+\infty} I(\lambda') \times \text{ILS}(\lambda, \lambda') d\lambda' \quad (3-37)$$

The ILS is a varying function of spectral color (pixel) within each footprint and band. Therefore, OCO-2 has  $1016 \times 8 \times 3 = 24384$  different line shape functions. Note that for GOSAT data based on an FTS, this step requires a convolution of the ILS with the monochromatic spectrum.

#### 3.2.4.2 Radiance Scaling

The calculated spectrum can be scaled by a detector pixel-dependent scale factor. For downlooking spectra, radiance scaling is not used. For uplooking spectra, where there is no analog to the scaling that occurs through ground model, this radiance scaling can be used to match the absolute values of the measured and modeled spectrum:

$$C(\lambda) = \sum_{n=1}^N c_k^n \times (\lambda - \lambda_c)^{n-1} \quad (3-38)$$

where  $c$  are  $N$  radiance scaling coefficients for spectrometer  $k$ , and  $\lambda_c$  is the center wavelength of the spectral window.

### 3.2.4.3 Instrument Doppler Correction

The wavelength scale is modified based on the relative velocity of the spacecraft and the Earth target to account for the Doppler shift. This relative velocity is calculated from the spacecraft velocity and the Earth's rotation. The equation for the Doppler correction is:

$$\lambda_{\text{adj}} = \lambda_{\text{org}} \left( 1 + \frac{v_{\text{rel}}}{c} \right) \quad (3-39)$$

where  $c$  is the speed of light. The L2 retrieval is then performed on this adjusted wavelength grid.

### 3.2.4.4 Zero-Level Offset

As discussed in Butz et al. [2011] and Crisp et al. [2012], Band 1 of the TANSO-FTS instrument on GOSAT is known to have a nonlinear response to intensity. This leads to systematic features in the Band 1 spectra, the size of which are intensity-dependent. To account for this, we include a wavelength-independent zero-level-offset, originally shown by Butz et al. [2011] to partially mitigate this instrument problem. This is modeled simply as a constant radiance added to the spectrum at TOA. It is given a prior value of zero with a very loose prior constraint.

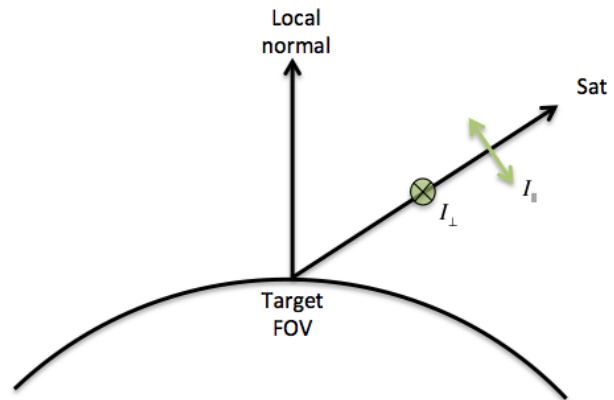
For OCO-2, we do not expect to need to retrieve a zero-level-offset for any band, though the capability is there if it proves to be necessary.

### 3.2.4.5 Treatment of Polarization

OCO-2 and OCO-3 fundamentally measure a single linear polarization of the incoming light. The axis of polarization accepted by the instrument depends on various instrument and spacecraft geometry parameters. For detailed description of these parameters, refer to the L1B ATBD. The important definition relevant for the L2 algorithm, is that for both OCO-2 and OCO-3 the polarization angle and stokes coefficients are defined using the meridian plane as the polarization reference plane. The meridian plane contains the local normal unit vector and the vector pointing from the target field of view (FOV) to the satellite (Figure 3-14). Both the polarization angle and the stokes coefficients are given in the L1B file.

### 3.2.4.6 Application of Stokes Vector

As described in the previous section, for OCO-2 and OCO-3 the only parameter necessary to calculate the Stokes coefficients is the angle of the slit with respect to the reference plane. These are calculated and stored in the L1B files. Within the L2 RT code, we simply calculate  $I$ ,  $Q$ , and  $U$  at the top of the atmosphere, and construct the measured radiance as in Equation 3-40. Specifically, we write this as:



**Figure 3-14.** Geometry of the reference plane for polarization, i.e., the local meridian plane. This plane is formed by the local normal unit vector and the ray from the target FOV to the satellite. The direction of propagation of measured light is the latter. Two example polarizations are shown depicting light polarized both parallel ( $\parallel$ ) and perpendicular ( $\perp$ ) to this plane. The polarization angle of the instrument is defined as the angle between the axis of accepted polarization and the reference plane for polarization.



$$I_{meas} = m_1 I + m_2 Q + m_3 U \quad (3-40)$$

where the  $m_i$  are the Stokes coefficients, and  $V$  is not calculated by the RT software (it is assumed to be negligible). In addition, the corresponding Stokes coefficient ( $m_4$ ) is equal to zero for OCO-2 and OCO-3. Equation (3-42) is performed for each channel and band on OCO-2 and OCO-3, using the per-band Stokes coefficients in the L1B FootprintGeometry group. The same equation is used for GOSAT, but the construction of the GOSAT Stokes parameters is significantly more complicated and is described in detail in O'Brien et al. [2013].

### 3.3 State Vector

#### 3.3.1 Basic Description of State Vector

The section below describes the state vector and setup used in the at-launch version of the L2 algorithm (B4.0). A previous version, B2.9, was described fully in the literature [O'Dell et al., 2012]. Below we primarily describe updates to that algorithm version.

A state vector specifies those aspects of the state of the atmosphere being measured and of the instrument measuring it, which together determine the value of the resulting measurement. We will use the term *state structure* to indicate the complete set of parameters required by the forward model to simulate a measurement to the necessary accuracy, and the term *state vector* to indicate the set of parameters being retrieved. This may be a simple subset of the state parameters that describe changes to the state structure (such as an offset to a temperature profile).

The state structure contains all the information about the atmosphere, the surface, the instrument, and the Sun needed by the forward model, as described in Section 3.2. These are captured in six substructures:

- Atmosphere: temperature profile, water vapor profile, and surface pressure
- Absorber: gaseous volume mix ratio profiles (CO<sub>2</sub>, H<sub>2</sub>O, O<sub>2</sub>, [O<sub>3</sub>, and CH<sub>4</sub>—not currently used])
- Aerosol: aerosol shape parameters (aerosol optical depth [AOD] at 755 nm, peak height, and width)
- Ground: Bidirectional reflectance distribution function parameters—either BRDF amplitude (mean and spectral slope), ocean BRDF parameters, including a lambertian mean and spectral slope, as well as the surface wind speed.
- Instrument: spectral dispersion and ILS
- Solar: parameters of solar continuum model
- Empirical orthogonal function (EOF) amplitudes for residual fitting

The state structure is completely flexible and can include an arbitrary number of absorber and aerosol extinction profiles with an arbitrary number of vertical levels. For operational L2 retrievals, 20 levels are used in the state structure for CO<sub>2</sub>, but 91 levels are used for temperature and water vapor. These are discussed below in Section 3.3.2.

The state vector currently consists of 60 or 64 elements (OCO-2 or OCO-3) or 52–53 elements (GOSAT) as shown in **Table 3-4**. The exact number depends on whether the retrieval is over land or ocean, and is slightly different for OCO-2 versus GOSAT.

**Table 3-4.** Description of state vector elements.

Description	Parameters	Number of Elements
Aerosols	5 x 3 parameters	15
Temperature offset	Scalar	1
Water vapor multiplier	Scalar	1
Surface pressure	Scalar	1
BRDF amplitude, spectral slope and quadratic curvature (land)	3 x 3 bands	9
Lambertian albedo mean & spectral slope (ocean)	2 x 3 bands	6
Surface wind speed (ocean only)	Scalar	1
CO <sub>2</sub>	20 levels	20
Spectral dispersion offset	1 shift per band	3
Spectral dispersion slope (OCO-2 only)	1 slope per band	3
O <sub>2</sub> A band radiance offset (GOSAT only)	Scalar	1
Chlorophyll fluorescence (O <sub>2</sub> A band) (land only)	Mean + slope	2
Residual EOF amplitudes	1 per band (GOSAT) 3 per band (OCO-2)	3 (GOSAT) 9 (OCO-2)
<b>OCO-2 total land (ocean)</b>		<b>64 (60)</b>
<b>GOSAT total land (ocean)</b>		<b>53 (52)</b>

In the next section, we describe the initial value and associated uncertainties of each state vector quantity.

### 3.3.2 The *a Priori* State Vector and Covariance Matrix

One of the most important factors in any optimal estimation retrieval algorithm is the setup of the *a priori* values and their corresponding covariance matrices. They represent our best guess of the values of the parameters given our prior knowledge, along with an estimate of their 1-sigma uncertainties (assuming Gaussian statistics). When the *a priori* constraint on a parameter is relatively loose (large *a priori* covariance), the retrieval relies more on the measurement in order to determine the parameter; conversely, when the *a priori* constraint is tight (small *a priori* covariance), the measurements have less effect on the retrieved parameter value. The overall *a priori* covariance matrix is the matrix of error variances and covariances of the *a priori* values of all state vector parameters; in principle, the *a priori* errors of any two parameters may be correlated.

The retrieval algorithm relies in part on constraining solutions so that they are statistically reasonable. In order to do this, we must have a realistic *a priori* state vector and a covariance matrix which constrains poorly measured components of the state. Note that, for convenience, the *a priori* state vector is also used for the first guess in our retrieval, as it ideally represents our best guess on the state of the atmosphere and surface before taking the OCO-2 observations into account.

The *a priori* vector is a simple vector of parameters with their *a priori* values. We take the corresponding *a priori* covariance matrix to be diagonal, with the exception of the carbon dioxide profile block. That is, the *a priori* uncertainty of every parameter is assumed to be uncorrelated with that of every other parameter, except within the *a priori* CO<sub>2</sub> profile, in which uncertainties at one vertical level may be correlated with uncertainties at other levels. We now describe the *a priori* values and assumed uncertainties for each parameter in turn.

### 3.3.2.1 Profile of Carbon Dioxide Concentration

The CO<sub>2</sub> *a priori* profiles are nearly identical to those used in the

GGG2020 TCCON retrieval. The priors are generated by using NOAA CO<sub>2</sub> monthly average flask measurements [Dlugokencky et al. 2019] averaged between the Mauna Loa and American Samoa sites (referred to below as MLO/SMO). These are assumed to represent the tropical average CO<sub>2</sub> concentration. The algorithm assumes that the CO<sub>2</sub> concentration at a given time, latitude, and altitude is a function of the MLO/SMO CO<sub>2</sub> concentration and time for transport between MLO/SMO and the latitude/altitude of the prior.

In the troposphere, this transport lag is computed using the same empirical function for age-of-air as used in the TCCON GGG2014 and GGG2020 retrievals. It requires as inputs the latitude of the sounding and altitude of the level in the prior, as well as the tropopause altitude for that sounding. The tropopause altitude is calculated from the blended tropopause pressure value in the GEOS-FPIT meteorology by interpolating the altitude grid to the tropopause pressure. With this age-of-air, the tropospheric CO<sub>2</sub> prior is calculated by looking up the MLO/SMO CO<sub>2</sub> concentration at the date the age-of-air indicates that MLO/SMO would have measured the CO<sub>2</sub> concentration for each tropospheric level of the prior. Finally, a correction is applied to account for the difference in latitudinal gradient in the priors compared to observations taken during the ATom campaigns. This bias is correlated with the age-of-air, therefore it can be removed by adding an offset as a function of CO<sub>2</sub> growth rate and age-of-air.

**Note:** the sole difference between the TCCON GGG2020 and OCO CO<sub>2</sub> profiles is in the latitude used in the age-of-air calculation. In GGG2020, this is an “effective latitude” computed from mid-tropospheric potential temperature; however, in the OCO product, the sounding latitude is used directly.

In the stratosphere, a similar process is followed, except that the age-of-air is taken from a climatology created with the Chemical Lagrangian Model of the Stratosphere (CLaMS) [McKenna et al. 2002] scaled to match the GSFC 2D model. This age is a function of day-of-

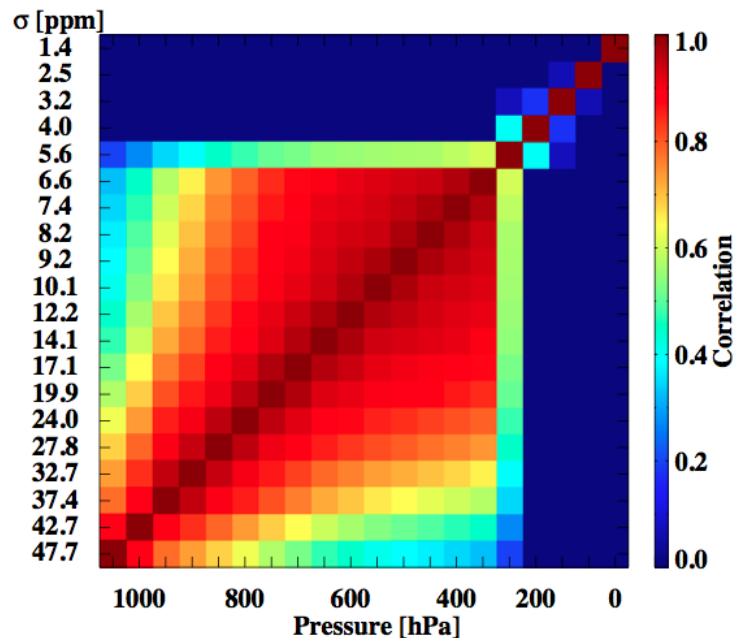


Figure 3-15. *A priori* CO<sub>2</sub> correlation matrix (in colors), together with the 1-sigma errors for each pressure level.

year, potential temperature, and potential vorticity-derived equivalent latitude. Additionally, to account for mixing of air of different ages during convective transport in the stratosphere, the MLO/SMO record is convolved with “age spectra” which are the analytical solutions to Green’s function [Andrews et al. 2001].

The tropospheric profile covers levels below the GEOS-FPIT blended tropopause pressure and the stratospheric profile covers levels above 380 K potential temperature, where isentropes lie entirely in the stratosphere. In between, the CO<sub>2</sub> profile is linearly interpolated with respect to potential temperature between the top tropospheric and bottom stratospheric level. The priors are initially calculated on the native 72 level GEOS-FPIT grid, then interpolated to the reduced grid used in the forward model. A detailed description of the priors will be available in a forthcoming publication [Laughner et al., in prep].

### **3.3.2.2 Meteorological Variables: Surface Pressure, Temperature, and Water Vapor**

Prior and first-guess meteorological variables are taken from a short-term forecast from the Goddard Space Flight Center (GSFC) Global Modeling and Assimilation Office (GMAO) Global Earth Observation System Forward Processing for Instrument Teams (GEOS FP-IT) product. The forecast length is between zero and nine hours. Details of the GEOS FP-IT product can be found at [https://gmao.gsfc.nasa.gov/GMAO\\_products/](https://gmao.gsfc.nasa.gov/GMAO_products/). The model includes 72 vertical levels and has a horizontal resolution of 0.3125° longitude by 0.25° latitude. GEOS FP-IT forecast files are downloaded daily.

Select fields are interpolated from the GEOS FP-IT files to the locations and times of the OCO-2 soundings, including surface pressure and profiles of temperature and specific humidity. Because the elevation of the model grid box is often slightly different from the actual elevation digital elevation map [Zong, 2008], a further correction is made based on the hydrostatic equation (often called a hypsometric adjustment) to account for this elevation difference.

The GEOS FP-IT surface pressures are assumed to have errors of approximately 1 – 2 hPa. It is unknown what the true error will be when over complex topography at scales of 1 – 2 km. Currently, the code assumes a 1-sigma error of 4 hPa on the *a priori* surface pressure estimate. This may be tightened in the future once the errors from OCO-2 are better understood. In contrast, for ACOS-GOSAT retrievals, a smaller prior error of 1 hPa is assumed.

As stated previously, the L2 FP code does not try to retrieve the entire temperature profile in the state vector. Instead, it is assumed that the GEOS FP-IT forecast temperature profile can be corrected by a single small offset. The *a priori* 1-sigma uncertainty in this offset is taken to be 5 K. It is likely that there is more information on temperature than just a single number, but thus far retrieval experiments indicate that this assumption for a temperature correction is sufficient.

Similarly to temperature, the water vapor profile from the GEOS FP-IT forecast is assumed to be correct up to an overall scale factor. This is consistent with the fact that there is not much water vapor information in the OCO-2 spectra, and most of it comes from the strong CO<sub>2</sub> band. Tests thus far indicate that this assumption is justified. We assume a 1-sigma error of 0.5 on the water vapor scale factor, which emphasizes the measurement in the retrieval.

### **3.3.2.3 Aerosol Profiles**

As stated previously, the FP model attempts to retrieve profiles of five scatterers: two cloud types, two tropospheric aerosol types, and one stratospheric aerosol type. The cloud types are a water cloud type and an ice cloud type. The two tropospheric aerosol types are drawn from dust, smoke, sulfate aerosol, organic carbon, and black carbon. In version 8 of the full-physics retrieval, a stratospheric aerosol was added to the retrieval state vector, which, for convenience, uses the same optical properties as the tropospheric sulfate aerosol type. The optical properties

of these types were explained in Section 3.2.1.6 above. For each type, we retrieve three quantities, which characterize the amount and vertical profile of that aerosol. The vertical profiles are assumed to be Gaussian in the relative pressure  $x = P/P_s$ , where  $P$  is the local pressure and  $P_s$  is the surface pressure. The shape of the Gaussian is defined by the location of the peak height and the 1-sigma width (both are defined in units of  $x = P/P_s$ ). The optical depth per unit pressure is given by:

$$\frac{d\tau_\lambda}{dx} = \frac{Q_\lambda}{Q_{755}} \frac{AOD_{755} S(x)}{\int_0^1 S(x) dx} \quad (3-41)$$

where  $Q_\lambda$  is the extinction efficiency at a wavelength  $\lambda$ ,  $AOD_{755}$  is the desired AOD at 755 nm (our reference wavelength), and  $S(x)$  is the profile shape given by:

$$S(x) = \exp\left\{-\frac{(x - x_a)^2}{2\sigma_h^2}\right\} \quad (3-42)$$

where  $x_a$  is the peak height of aerosol type  $a$ , and  $\sigma_a$  is the  $1\sigma$  Gaussian width. The *a priori* values and uncertainties are given in Table 3-5 below. Currently, the aerosol *a priori* error covariance matrix is strictly diagonal. The *a priori* uncertainty on the two tropospheric aerosol types has been reduced from 2.0 (on  $\ln(AOD)$ ) to 0.5, in accordance with Nelson et al. [2019] who demonstrated a small improvement in  $X_{CO_2}$  errors when using a lower *a priori* uncertainty coupled with better informed *a priori* AODs from GEOS-5. Prior aerosol profiles are shown in Figure 3-16.

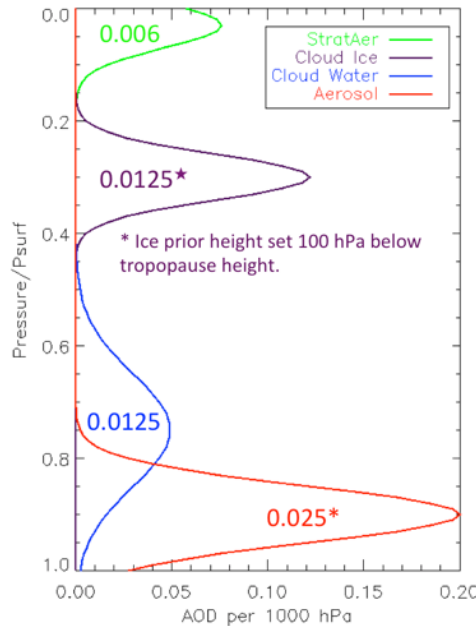


Figure 3-16. Prior aerosol profiles uses the L2FP retrieval.

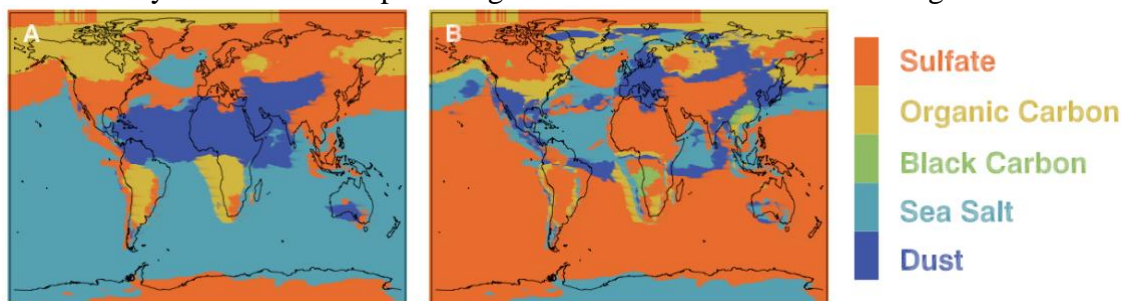
**Table 3-5.** Aerosol prior profile quantities and *a priori* uncertainties.

Type	AOD <sub>755</sub>	Ln(AOD) $\pm 1\sigma$	Height $\pm 1\sigma$	Width $\pm 1\sigma$
Water cloud	0.0125	$-4.382 \pm 1.8$	$0.75 \pm 0.4$	$0.1 \pm 0.01$
Ice cloud	0.0125	$-4.382 \pm 1.8$	variable <sup>†</sup> $\pm 0.2$	$0.04 \pm 0.01$
Aerosol 1	Variable*	variable* $\pm 0.5$	$0.9 \pm 0.2$	$0.05 \pm 0.01$
Aerosol 2	Variable*	variable* $\pm 0.5$	$0.9 \pm 0.2$	$0.05 \pm 0.01$
Stratospheric Aerosol	0.006	$-5.116 \pm 1.8$	$0.03 \pm 1e-4$	$0.04 \pm 0.01$

\*AOD priors for aerosol types are chosen from GEOS-5.

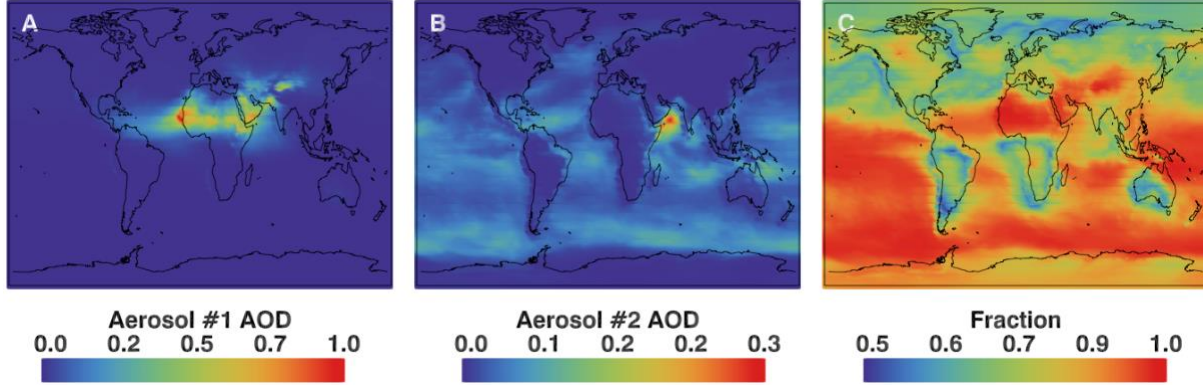
<sup>†</sup> The ice cloud height is set to be 100 hPa below the tropopause pressure (see Section 3.3.2.1)

The two aerosol types are chosen from co-located GEOS-5 FP-IT aerosols. They are the types that form the highest and second highest fraction of the AOD at 755 nm for a co-located GEOS-5 grid cell for a given sounding. Figure 3-17 shows maps of the dominant and second-most dominant type for MERRA data for July of 2009. Figure 3-18 shows the corresponding AODs for each of these types. Rather than use the peak height from GEOS-5, we have found it more numerically stable to set the peak heights and widths to fixed values as given in Table 3-5.



**Figure 3-17.** Global distribution of modeled aerosol type for July 2009. (**Panel A**) dominant type; (**Panel B**) second-most dominant type.

Using modeled heights from GEOS-5 proved ineffective, as models struggle to realistically distribute aerosols vertically in the atmosphere [Kipling et al., 2016]. For all scatterers, the peak width prior uncertainty is set so low (0.01) that it is effectively not retrieved. This is because in practice our spectra have very little sensitivity to the width of a thin scattering layer.



**Figure 3-18.** Global distribution of modeled aerosol optical depth for July 2009. (**Panel A**) dominant type; (**Panel B**) second-most dominant type; (**Panel C**) fractional contribution of the two dominant types to total AODs.

### 3.3.2.4 Surface Properties

#### 3.3.2.4.1 Land Surface *a Priori*

##### Definition of the BRDF model:

A general BRDF model can be made up of two kernels, for the unpolarized and polarized components, respectively, of the reflection. For azimuthally symmetric BRDF models, this can be expressed as:

$$\rho(\theta_i, \theta_r, \Delta\varphi) = r[r_s F_s(\theta_i, \theta_r, \varphi) + r_p F_p(\theta_i, \theta_r, \varphi)] \quad (3-43)$$

where:

$\rho$  is the BRDF

$r$  is the overall amplitude

$r_s$  is the amplitude of the scalar (RPV) model

$F_s$  is the shape of the scalar (RPV) model

$r_p$  is the amplitude of the polarized (Bréon) model

$F_p$  is the shape of the polarized (Bréon) model

$\theta_i$  is the incident zenith angle

$\theta_r$  is the reflection zenith angle

$\varphi$  is the relative azimuth angle

The BRDF models are functions of three angles, under the assumption that the surface is azimuthally symmetric. Thus, only the relative azimuth angle (difference between solar and observer azimuth) is needed.

For the OCO-2 forward model, we assume that  $r_p = 0$  and  $r_s = 1.0$ , making the BRDF unpolarized, and only dependent on the RPV (Rahman – Pinty – Verstraete) [Rahman et al. 1993]. Dropping the polarized kernel, we can write the BRDF as:

$$\rho(\theta_i, \theta_r, \varphi) = r r_s F_s(\theta_i, \theta_r, \varphi; \rho_0, \Theta, k) \quad (3-44)$$



where now the BRDF shape is described by certain kernel parameters:  $\rho_0$ ,  $\Theta$ ,  $k$ . The forward model takes fixed values for these parameters, which define the actual shape of the RPV kernel.

Further,  $r$  can be defined as follows, to introduce a quadratic relationship of the BRDF amplitude with wavenumber.

$$r = w + s(\nu - \nu_0) + q(\nu - \nu_0)^2 \quad (3-45)$$

where  $w$  is the “BRDF weight”,  $s$  is the “BRDF weight slope”, and  $q$  is the “BRDF quadratic” term. For a Lambertian model, these would be the albedo, spectral slope, and quadratic curvature term, respectively.  $\nu$  and  $\nu_0$  are the wavenumber and reference wavenumber (at which the surface properties are retrieved), respectively.

We can therefore rewrite the BRDF in the following manner, to include the wavenumber dependence:

$$\rho(\theta_i, \theta_r, \varphi, \nu) = \rho_w(\theta_i, \theta_r, \varphi) + \rho_s(\theta_i, \theta_r, \varphi)(\nu - \nu_0) + \rho_q(\theta_i, \theta_r, \varphi)(\nu - \nu_0)^2 \quad (3-46)$$

where:

$$\rho_w(\theta_i, \theta_r, \varphi) = wr_s F_s(\theta_i, \theta_r, \varphi; \rho_0, \Theta, k) \quad (3-47)$$

$$\rho_s(\theta_i, \theta_r, \varphi) = sr_s F_s(\theta_i, \theta_r, \varphi; \rho_0, \Theta, k) \quad (3-48)$$

$$\rho_q(\theta_i, \theta_r, \varphi) = qr_s F_s(\theta_i, \theta_r, \varphi; \rho_0, \Theta, k) \quad (3-49)$$

The OCO-2 L2 full physics retrieval labels the variables as follows. First, the three retrieved state variables:

$w$  : brdf\_weight

$s$  : brdf\_weight\_slope

$q$  : brdf\_weight\_quadratic

and then the 5 fixed state parameters:

$r_s$  : brdf\_rahman\_factor = 1.0

$r_p$  : brdf\_breon\_factor = 1e-20 (effectively zero)

$\rho_0$  : brdf\_hotspot\_parameter = 0.05

$\Theta$  : brdf\_asymmetry\_parameter = -0.1

$k$  : brdf\_anisotropy\_parameter = 0.75

These values for the fixed BRDF parameters were used as the initial guess for an inversion algorithm using multi-angle polarimetric satellite data (Table 1 and 6 in Dubovik et al. [2011]).

Note that the three state variables are retrieved independently within each band, but all bands use the same values of the fixed parameters. The a priori values for the brdf\_weights are set according to a simple conversion of each band’s observed continuum radiance. The radiance is converted to a top-of-atmosphere (TOA) reflectance with the following expression:

$$\rho_{\text{TOA}} = \frac{\pi L_{\text{cont}}}{m_1 S \cos \theta_i} \quad (3-50)$$

where  $L_{\text{cont}}$  is the continuum radiance,  $m_1$  is the first stokes coefficient (see section 3.2.4.6),  $S$  is the distance-corrected solar irradiance, and  $\theta_i$  is the solar zenith angle at the observed ground footprint. The TOA reflectance is then converted into the equivalent brdf\_weight by computing the amplitude that produces that reflectance value for the given sounding geometry. The slope and quadratic terms use a priori values of zero. All BRDF parameters are set with very loose a priori constraints. Table 3-6 shows the values used for the a priori mean and covariance.

### Output Variables:

The three retrieved state variables ( $w$ ,  $s$ , and  $q$ ; the brdf\_weight, brdf\_weight\_slope, and brdf\_weight\_quadratic, respectively) are converted to more physically meaningful surface reflectance values (brdf\_reflectance, brdf\_reflectance\_slope, and brdf\_reflectance\_quadratic) for the L2 retrieval output. The reflectance values are computed by evaluating the weighted BRDF kernel model at the primary observation geometry (solar incident angle, reflection angle to the sensor). The result is a reflectance factor, equal to the ratio between the surface reflected radiance and the radiance from a purely reflecting Lambertian surface at the same geometry [Schaeppman-Strub et al. 2006]. For conditions with no atmospheric scattering, this should be equivalent to a Lambertian albedo.

Evaluating the RPV kernel at the primary angle can be expressed as follows, where the subscript 0 values denote the solar and observer angles:

$$\rho_w(\theta_i = \theta_{i,0}, \theta_r = \theta_{r,0}, \varphi = \varphi_0) = wr_s F_s(\theta_i = \theta_{i,0}, \theta_r = \theta_{r,0}, \varphi = \varphi_0; \rho_0, \Theta, k) \quad (3-51)$$

$$\rho_s(\theta_i = \theta_{i,0}, \theta_r = \theta_{r,0}, \varphi = \varphi_0) = sr_s F_s(\theta_i = \theta_{i,0}, \theta_r = \theta_{r,0}, \varphi = \varphi_0; \rho_0, \Theta, k) \quad (3-52)$$

$$\rho_q(\theta_i = \theta_{i,0}, \theta_r = \theta_{r,0}, \varphi = \varphi_0) = qr_s F_s(\theta_i = \theta_{i,0}, \theta_r = \theta_{r,0}, \varphi = \varphi_0; \rho_0, \Theta, k) \quad (3-53)$$

**Table 3-6.** Surface state variables and a priori uncertainties.

variable	band	a priori mean	uncertainty ( $\pm 1\sigma$ )
brdf_weight	All	Variable (see text)	5
brdf_weight_slope	O <sub>2</sub> A	0	0.001
brdf_weight_quadratic	O <sub>2</sub> A	0	5.0E-6
brdf_weight_slope	weak CO <sub>2</sub>	0	0.003
brdf_weight_quadratic	weak CO <sub>2</sub>	0	3.0E-5
brdf_weight_slope	strong CO <sub>2</sub>	0	0.005
brdf_weight_quadratic	strong CO <sub>2</sub>	0	1.0E-4
Wind speed	N/A	Variable (from meteorology)	6.325 m/s

variable	band	a priori mean	uncertainty ( $\pm 1\sigma$ )
Lambertian albedo	O <sub>2</sub> A	0.02	0.2
Lambertian albedo slope	O <sub>2</sub> A	0	1
Lambertian albedo	weak CO <sub>2</sub>	0.02	0.2
Lambertian albedo slope	weak CO <sub>2</sub>	0	1
Lambertian albedo	strong CO <sub>2</sub>	0.02	0.001
Lambertian albedo slope	strong CO <sub>2</sub>	0	1

### 3.3.2.4.2 Ocean Surface *a Priori* in Glint Mode

Over water surfaces (defined as land fraction less than 20%), the surface is assumed to be described by a combination of two surface types: A Cox-Munk distribution of planar facets [Cox and Munk, 1954] plus a Lambertian surface. The latter is necessary because while the Cox-Munk surface can usually describe our observations quite well, it cannot perfectly fit the continua in all three spectral bands with a single free parameter (the surface wind speed).

There are a number of options that could be used to fit the surface. One could (unphysically) fit a separate wind speed for each band, but this still leaves minor problems with the slope of the continuum in a given band being inconsistently fit. Alternatively, one could fit the water index of refraction (and its slope) in each spectral band. We choose to fit a low-reflectivity Lambertian surface on top of the Cox-Munk surface. The water Lambertian surface has a mean albedo and albedo slope separately specified for each band. The *a priori* mean values are set to 0.02 in the O<sub>2</sub> A, weak CO<sub>2</sub>, and strong CO<sub>2</sub> bands. We constrain the strong CO<sub>2</sub> mean albedo so it cannot take a value below 0.02, and we require non-negative Lambertian albedos in each band. The mean albedos in the other two bands have *a priori* uncertainties of 0.2.

For the Cox-Munk surface, the refractive index is taken to be constant within a band, with values of 1.331, 1.318, and 1.303 in the O<sub>2</sub> A, weak CO<sub>2</sub>, and strong CO<sub>2</sub> bands, respectively. The *a priori* wind speed is taken from the GEOS FP-IT forecast, with a prior 1-sigma uncertainty of 6.325 m/s.

### 3.3.2.4.3 Chlorophyll Fluorescence

Over land surfaces, the state vector also contains the value of chlorophyll fluorescence at the surface at  $\lambda_{\text{ref}} = 757$  nm (or its equivalent in wavenumbers,  $\nu_{\text{ref}}$ ), plus its slope with respect to wavenumber. Our approach was introduced in Frankenberg et al. [2012], where it was shown that ignoring chlorophyll fluorescence in XCO<sub>2</sub> retrievals that use the O<sub>2</sub> A band would generally induce biases in the retrieved XCO<sub>2</sub>, regardless of whether the O<sub>2</sub> A band was only used to constrain aerosols, surface pressure, or both. Specifically, we assume that the SIF signal at the surface, SIF<sub>surf</sub>, can be adequately parameterized across the O<sub>2</sub> A band as:

$$SIF_{\text{surf}}(\nu) = f_{\nu_{\text{ref}}} (1 + \gamma [\nu - \nu_{\text{ref}}]) \quad (3-54)$$

where  $\nu$  is the wavenumber in units of  $[\text{cm}^{-1}]$ . The two state vector parameters are  $f_{\lambda_{\text{ref}}}$ , or equivalently,  $f_{\nu_{\text{ref}}}$ , the value of the surface-level fluorescence radiance (in units of  $[\text{ph/s/m}^2/\text{sr}/\mu\text{m}]$ ) at the reference wavelength (or wavenumber), and the slope term  $\gamma$  in units of

[1/cm<sup>-1</sup>]. Note that internally, the fluorescence spectrum is calculated on a spectral grid in wavenumber units, so the slope term must be defined in inverse wavenumbers. The *a priori* values and 1-sigma errors are set as:

$$f_{\lambda_{ref}} = \frac{IMAP\ SIF_{757nm} + 1.8 \cdot IMAP\ SIF_{771nm}}{2} \pm \sigma(IMAP\ SIF_{757nm}) \quad (3-55)$$

$$\gamma = 1.8 \cdot 10^{-3} \pm 7 \cdot 10^{-6}$$

The prior value for the fluorescence at the reference wavelength (or wavenumber) is constructed as a weighted average of the fluorescence obtained from dedicated single-band retrievals (IMAP) at the two wavelengths at 757nm and 771nm. The 1-sigma error is carried over from the single-band retrieval posterior uncertainty.

The fluorescence signal is modeled as isotropic and unpolarized. The RT model propagates the surface signal to TOA, as discussed in Section 3.2.3.4. It should be noted that the primary purpose of the fluorescence implementation here is to minimize biases in XCO<sub>2</sub> and not provide the best possible fluorescence product. For fluorescence data users, fits based purely on Fraunhofer lines will be more robust and will be distributed within the IMAP pre-processor.

### 3.3.2.5 Wavelength Grid Offset

Based on our experience with GOSAT retrievals, there could be a wavelength offset. We measured the wavelength grid of OCO-2 on the ground. This is a wavelength (in  $\mu\text{m}$ ) for each of the 1016 channels for each footprint and band. On orbit this may be slightly displaced due to instrumental and Doppler effects. This offset is fit separately for each spectral band, but a good first guess is necessary. If the offset is off by more than about one spectral sample, the fit will often fail to converge.

We will initially try using a simple instrument Doppler correction. If this correction is not enough, we will try and estimate the grid offset initial guess. We estimate this offset based on observing the actual position of a strong solar line near 12985.16326 cm<sup>-1</sup> and assuming that the Doppler multiplier is the same for each band. The first guess is typically good to approximately 0.02 cm<sup>-1</sup>, sufficient for a good first guess. The *a priori* uncertainty of this value is set relatively large, so in reality the L2 fit completely constrains this parameter (for each footprint and band).

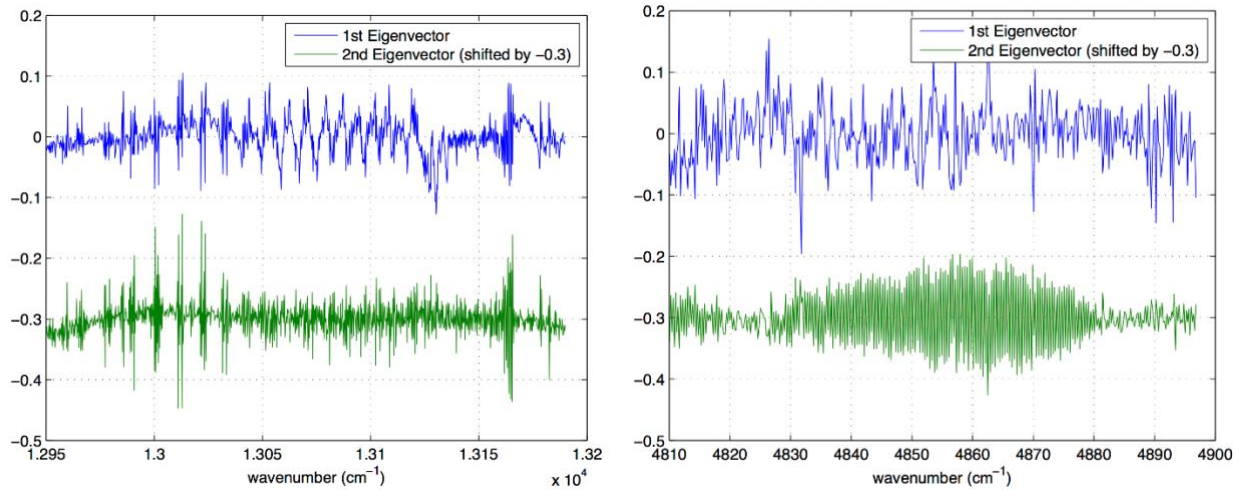
### 3.3.2.6 Residual EOF Parameters

*Note: This subsection describes a set of state vector elements that was introduced in Atmospheric CO<sub>2</sub> Observations from Space (ACOS) algorithm version B3.3. It is not described in either O'Dell et al. [2012] or Crisp et al. [2012]. It replaces the older “empirical noise” formulation used in previous versions of the algorithm.*

#### 3.3.2.6.1 Using Empirical Orthogonal Functions to Minimize Fit Residuals

Owing to remaining inaccuracies of the spectroscopic input data and also to instrumental effects, we almost never obtain residuals consistent with the detector noise estimates, especially at higher SNR. If systematic residuals persist, the reduced  $\chi^2$  will thus scale with SNR and reach 1 only at very low SNR values.  $\chi^2$  values would thus not truly represent the goodness of the fit anymore and make filtering based on these values difficult.

Based on experience with GOSAT retrievals, we adopted an alternative approach using a singular value decomposition of many fit residuals. The method works as follows: For a diverse training set of good fits (without compensating for systematic residuals), we compose three



**Figure 3-19.** Example of first two leading eigenvectors for current GOSAT retrievals in the O<sub>2</sub> A band (**left**) and strong CO<sub>2</sub> band (**right**).

matrices  $\mathbf{M}$  constituted by the spectral residuals of the spectral fits within each spectral band (typically >10000 soundings are used). Each matrix  $\mathbf{M}$  is then decomposed into its eigenvectors using singular value decomposition:

$$\mathbf{M} = \mathbf{U} \mathbf{S} \mathbf{V}^T \quad (3-56)$$

with the columns of  $\mathbf{U}$  spanning orthonormal basis vector of the systematic residuals observed in the training dataset. By convention, the first eigenvector explains the largest fraction of the total variance, as indicated by descending order of singular values (the diagonal elements of  $\mathbf{S}$ ). In most cases encountered here, the first singular vector will be very similar to the mean residual. However, the analysis of higher order singular vector can reveal other systematic features (be they instrumental or spectroscopic) and can sometimes contribute to the overall observed variance in residuals.

Examples of typical eigenvectors encountered in fit residuals with GOSAT data can be seen in Figure 3-19. In the case of the O<sub>2</sub> A band, these features are mostly related to errors in the oxygen spectroscopy, especially in terms of its line mixing behavior as well as collision-induced absorption and specific line shape model. The second eigenvector appears to be related to a change in line width, which might be an instrumental effect in this band (as the instrument line shape might vary if the field of view is not homogeneously illuminated). For the strong CO<sub>2</sub> band, the first eigenvector is again mostly related to spectroscopy and potentially also the broad-band structure of the solar irradiance. The second eigenvector is related to an instrumental artifact that has been eliminated in the latest GOSAT L1B version. These examples show what can be learned from the systematic analysis of spectral residuals, and we expect the leading eigenvector to be similar for OCO-2, unless major changes in the spectroscopy are achieved by that time.

### 3.3.2.6.2 Retrieval Implementation

The procedure for including singular vectors in the L2 FP retrievals works as follows. We know that the systematic features can be explained by a linear combination of eigenvectors as illustrated above. We can then modify our original forward modeled radiance values  $I_b(\mathbf{x})$ , which represent a vector of radiances in a given spectral band  $b$ , by adding a linear combination of all eigenvector  $\mathbf{u}_{i,b}$ , multiplied by a scalar  $a_i$ . Here  $i$  represents the eigenvector number. The factors

$a_{i,b}$  (for each band) are additional parts of the state vector to be retrieved. The corresponding Jacobians are simply the vectors  $\mathbf{u}_i$ . The modification to the modeled radiance is then simply:

$$\vec{I}'_b(\mathbf{x}) = \vec{I}_b(\mathbf{x}) + \sum_i a_i \vec{u}_i \quad (3-57)$$

For GOSAT, we used one eigenvector per band, and have achieved reduced  $\chi^2$  values close to 1 irrespective of signal level. For OCO-2, the number of eigenvectors used in the retrieval will depend on the fractional contribution to the overall variance, which will only be known once it is in orbit and in initial retrievals are made.

### 3.3.2.6.3 Advantages and Risks

One of the advantages of this method is that the goodness of the fit will be largely improved with better convergence. Theoretically, the inclusion of residual-pattern eigenvectors in the fit should be mostly cosmetic, but it can affect the weighting of different spectral regions and thereby also the retrieval. Especially broad-band systematic residuals that deviate from zero will be penalized strongly by the least-squares fit routine. The inclusion of the eigenvectors of the fit residuals ensures that most of the noise is weighted properly with the mean residual being close to flat.

The risk is that systematic problems may be hidden by the eigenvectors. Ideally, these features should not occur in the first place and our goal should be to minimize the need for eigenvectors as much as possible. Also, the generation of eigenvectors is somewhat circular. For every code or spectroscopy change, a large dataset has to be run without the EOF fit turned on to generate a training dataset for the computation of the updated eigenvectors.

## 3.4 Jacobians

In the inverse method we use, the entire  $K$ -matrix of the state vector is required, which represents the partial derivative of each spectral channel radiance with respect to each state vector element. Mathematically, this is expressed as:

$$K_{ij} = \frac{\partial F_i(\mathbf{x})}{\partial x_j} \quad (3-58)$$

where  $x_j$  is the state vector element  $j$  and  $F_i$  is the forward modeled radiance for channel  $i$ . The typical retrieval setup for OCO-2 has  $m = 3048$  channels and  $n = 45$  state vector elements, meaning that  $K$  is a (3048, 45) matrix.

There are two components to the calculation of the Jacobian. The LIDORT code we use is linearized: it can calculate the Jacobian of its output. This Jacobian is in terms of the inputs and outputs of LIDORT. The inputs to LIDORT are *not* the same as our state vector elements, although they are calculated from them. Likewise, the outputs from LIDORT are not the same as our forward model, although the forward model is calculated using LIDORT and the state vector elements.

We perform the mapping from state vector elements to the LIDORT input and the LIDORT output to the forward model automatically using the standard calculus chain rules implemented by automatic differentiation.

## Automatic Differentiation

We make use of a technique called Rall numbers, as described in *Scientific and Engineering C++* by Barton and Hackman [1997] (named for the original book by L.B. Rall *Automatic Differentiation*, Springer-Verlag, Berlin, New York, 1981). In our code, instead of using a standard C *double* we use a new C++ class *AutoDerivative* that keeps track of both a value and the gradient of the value with respect to the state vector elements. All the normal mathematical functions are overloaded to use the chain rule to propagate the gradient through the calculation.

One of the advantages of this particular technique is that it works well with code that calculates Jacobians in other ways. This is one of the primary reasons we selected this technique over some of the other automatic derivative techniques. For example, LIDORT is already calculating Jacobians—we have no interest in replacing anything in LIDORT. Instead, we calculate the Jacobian of the parameter up to LIDORT, use LIDORT itself to propagate through the RT calculation, and then take the results of LIDORT, wrap them up as *AutoDerivative* objects, and continue using this to propagate through the remaining forward model calculation.

Although the use of an *AutoDerivative* is computationally fast, it is not as fast as hand-coded derivative calculations. For locations in the code that profiling show as bottlenecks, we can replace the *AutoDerivative* calculation with hand-created code that gives the same results.

## 3.5 Inverse Method

### 3.5.1 Formulation and Implementation

As mentioned in Section 3.1, the notation and concepts of Rodgers [2000] are employed for the inverse method. This section builds on Section 2.1 of Connor et al. [2008]. The spectrum, or measurement vector  $\mathbf{y}$ , is expressed symbolically as  $\mathbf{y} = \mathbf{F}(\mathbf{x}) + \boldsymbol{\varepsilon}$  where  $\mathbf{x}$  is the state vector,  $\mathbf{F}$  is the forward model, and  $\boldsymbol{\varepsilon}$  is the vector of spectral errors due to the measurement and forward model.

The solution of the OCO-2 inverse method is the state vector  $\hat{\mathbf{x}}$  with maximum *a posteriori* probability, given the measurement  $\mathbf{y}$ . Our inverse method employs the Levenberg-Marquardt modification of the Gauss-Newton method.

The operational inverse method consists of a set of routines, which are essentially mathematical and independent of the physics embodied by the measurement and state vectors. This implies that the structure of both vectors may be varied, so the routines are readily applied to other experiments, such as the Scanning Imaging Absorption Spectrometer for Atmospheric Chartography or ground-based FTS [Boesch et al., 2006]. The ability to retrieve  $X_{\text{CO}_2}$  from space-based and ground-based measurements using the same algorithm is critical for detecting and removing biases from the space-based data and forms a critical component of the OCO-2 validation strategy [Crisp et al., 2004].

The result of the inverse method is classified according to the parameter *outcome* as described in Table 3-7.

**Table 3-7.** Description of *outcome* values.

<i>outcome</i>	Description
1	Convergence was reached and the spectral fit was consistent with assumed uncertainties
2	Convergence was reached but the spectral fit was poor relative to assumed uncertainties
3	Convergence was not reached in the maximum allowed number of iterations
4	The maximum number of diverging steps (described below) was exceeded



To find the state vector that produces the maximum *a posteriori* probability, we minimize the following standard cost function  $c$ :

$$\underline{c} = (\mathbf{y} - \mathbf{F}(\mathbf{x}))^T \mathbf{S}_\varepsilon^{-1} (\mathbf{y} - \mathbf{F}(\mathbf{x})) + (\underline{\mathbf{x}}_a - \mathbf{x})^T \mathbf{S}_a^{-1} (\underline{\mathbf{x}}_a - \mathbf{x}) \quad (3-59)$$

where the weighting function matrix (Jacobian) is:

$$\mathbf{K} = \frac{\partial \mathbf{F}(\mathbf{x})}{\partial \mathbf{x}} \quad (3-60)$$

and  $\underline{\mathbf{x}}_a$  is the *a priori* state vector,  $\mathbf{S}_a$  is the *a priori* covariance matrix, and  $\mathbf{S}_\varepsilon$  is the measurement covariance matrix. On each iteration  $i$ , we solve for the state vector update  $\mathbf{dx}_{i+1}$  using a slightly modified form of Rodgers' Equation 5.36 to improve numerical accuracy by avoiding inversion of a large matrix:

$$((1 + \gamma)\mathbf{S}_a^{-1} + \mathbf{K}_i^T \mathbf{S}_\varepsilon^{-1} \mathbf{K}_i) \mathbf{dx}_{i+1} = [\mathbf{K}_i^T \mathbf{S}_\varepsilon^{-1} (\mathbf{y} - \mathbf{F}(\mathbf{x}_i)) + \mathbf{S}_a^{-1} (\underline{\mathbf{x}}_a - \mathbf{x}_i)] \quad (3-61)$$

where  $\gamma$  is the Levenberg-Marquardt parameter. This method reduces to regular Gauss-Newton minimization when  $\gamma = 0$ .  $\gamma$  is initialized with a value of 10.0.

After each calculation of  $\mathbf{dx}_{i+1}$ , before using it to update  $\mathbf{x}_i$ , we assess the impact of forward model nonlinearity on the reduction in the cost function. To do this, we calculate the following quantity:

$$R = (\mathbf{c}_i - \mathbf{c}_{i+1}) / (\mathbf{c}_i - \mathbf{c}_{i+1,FC}) \quad (3-62)$$

Here,  $c_i$  and  $c_{i+1}$  are the cost function from Equation (3-59) after iterations  $i$  and  $i+1$ , and  $c_{i+1,FC}$  is calculated from Equation (3-59) assuming that  $F(\mathbf{x} + \mathbf{dx}_{i+1}) = F(\mathbf{x}) + \mathbf{K} \mathbf{dx}_{i+1}$

This is simply the ratio of the actual reduction in the cost function  $c$  (between iterations  $i$  and  $i+1$ ) to the forecast reduction in the cost function based on an assumption of linearity. For a perfectly linear forward model, the cost function reduction will equal the forecast cost function reduction, and  $R$  will equal unity. A step in which the cost function actually increases will yield  $R < 0$ . Cases for which  $0 < R < 1$  are cases for which the cost function decreased, but not by as much as expected.

The use of  $R$  to assess  $\mathbf{dx}_{i+1}$  and to change  $\gamma$  is an area of active research and may be modified in future versions of the algorithm. In the algorithm immediately post launch, the detailed treatment is as described in the following paragraph.

Cases with  $R \leq 0.0001$  are considered divergent.  $\gamma$  is increased by a factor 10 and a new value of  $\mathbf{dx}_{i+1}$  is calculated. If the total number of diverging steps passes a certain threshold, the entire retrieval is abandoned, and *outcome* is set to 4. In cases with  $0.0001 < R$ ,  $\mathbf{dx}_{i+1}$  is used to update  $\mathbf{x}_i$ . If  $0.0001 < R < 0.25$ ,  $\gamma$  is increased by a factor 10 for the next iteration. For steps with  $0.25 < R < 0.75$ ,  $\gamma$  is left unchanged. Those cases with  $R > 0.75$  are deemed quite linear, and in these cases the Levenberg-Marquardt parameter  $\gamma$  is reduced by a factor of two.

After each successful, non-divergent iteration, we again test for convergence. To facilitate that, we compute the error variance derivative:

$$d\sigma_i^2 = d\mathbf{x}_{i+1}^T \hat{\mathbf{S}}^{-1} d\mathbf{x}_{i+1} \quad (3-63)$$

where  $\hat{\mathbf{S}}$  denotes the covariance of the error in the retrieved state using the relation:

$$\hat{\mathbf{S}}^{-1} d\mathbf{x}_{i+1} \cong [\mathbf{K}_i^T \mathbf{S}_\varepsilon^{-1} (\mathbf{y} - F(\mathbf{x}_i)) + \mathbf{S}_a^{-1} (\mathbf{x}_i - \mathbf{x}_a)] \quad (3-64)$$

and  $d\mathbf{x}_{i+1}$  is the state vector update assuming  $\gamma = 0$  (this is generally larger than the actual step size  $d\mathbf{x}_{i+1}$  computed for nonzero  $\gamma$ ).  $d\sigma^2$  is effectively the square of the state vector update in units of the solution variance.

If  $d\sigma^2 < fn$  (where  $n$  is the number of state vector elements, and  $f \sim 1$ ), convergence is reached. We then update the state vector a final time.

Lastly, we compute the retrieval covariance matrix  $\hat{\mathbf{S}}$  and the averaging kernel matrix  $\mathbf{A}$ .  $\hat{\mathbf{S}}$  is given by:

$$\hat{\mathbf{S}} = (\mathbf{K}^T \mathbf{S}_\varepsilon^{-1} \mathbf{K} + \mathbf{S}_a^{-1})^{-1} \quad (3-65)$$

and the averaging kernel matrix  $\mathbf{A}$  is given by:

$$\mathbf{A} = \hat{\mathbf{S}} \mathbf{K}^T \mathbf{S}_\varepsilon^{-1} \mathbf{K} \quad (3-66)$$

Finally, the degrees of freedom for signal are given by the trace of the matrix  $\mathbf{A}$ ; the degrees of freedom for the CO<sub>2</sub> profile are the trace of the CO<sub>2</sub>-only submatrix.

We note that the OCO-2 retrieval problem as described is underdetermined. There are typically  $\leq 2$  degrees of freedom for the CO<sub>2</sub> profile (compared to 20 profile levels) and  $\sim 20$  total degrees of freedom (compared to  $\sim 50$  elements). However, the use of an *a priori* constraint guarantees that the problem is well posed and well conditioned. Although there are no more than two degrees of freedom for the CO<sub>2</sub> profile (and thus there is little ability to discriminate between altitudes), the profile, rather than just the CO<sub>2</sub> column, is formally retrieved. This is to allow for the large variations of CO<sub>2</sub> in the boundary layer; such variations can produce bias in the results if a fixed profile shape is used in the retrieval.

### 3.5.2 Goodness of Spectral Fit

The goodness of fit is estimated from  $\chi^2$  for each spectral band:

$$\chi_i^2 = \frac{\frac{1}{n} \sum (\mathbf{y} - F(\mathbf{x}))^2}{\frac{1}{n} \sum \varepsilon^2} \quad (3-67)$$

If  $\chi_i^2 < \text{max\_chi2}$  for all  $i$ , then the fit is provisionally acceptable, and *outcome* is set to 1; otherwise, *outcome* = 2.

Because of the presence of non-random spectral errors, the values of *max\_chi2* must be determined empirically by inspection of a large set of retrieved results. Thus, they (and hence the assessment of fit quality) are subject to refinement. Therefore users may wish to include retrievals with both *outcome* = 1 and *outcome* = 2 in their initial analyses.

### 3.6 Xco2, Characterization, and Error Analysis

#### 3.6.1 Pressure Weighting Function

We define the pressure weighting function  $\mathbf{h}$  to relate the local CO<sub>2</sub> mixing ratio values specified on the discrete pressure levels to the profile-weighted average such that  $X_{CO_2} = \mathbf{h}^T \hat{\mathbf{x}}$ . This procedure is described in detail in O'Dell et al. [2012]. We only briefly summarize the procedure here—the interested reader is referred to that work for details. Recalling that  $X_{CO_2}$  is defined as the CO<sub>2</sub> column divided by the dry air column, it may be seen that the vector  $\mathbf{h}$  represents the pressure intervals assigned to the state vector levels, normalized by the surface pressure and corrected for the presence of water vapor such that  $\hat{\mathbf{x}}$  is weighted by the number of moles of dry air per layer.

First, we describe our pressure-leveling scheme. This is a simple sigma coordinate system where:

$$p_i = b_i p_{surf} \quad (3-68)$$

where the  $b_i$  constants are fixed. Currently, we have  $N_{lev}=20$  levels, and  $b = \{0.0001, 1/19, 2/19, \dots, 18/19, 1.0\}$ . Thus, the lowest level is equal to the surface pressure (this is slightly easier to work with than a previous approach we used, wherein the pressure levels were fixed). For  $p_{surf} = 1000$  hPa, the pressure width of all the layers (except the top layer) is 52.63 hPa; for the top layer it is 52.53 hPa. The typical model lid is therefore roughly 0.1 hPa.

In order to calculate  $\mathbf{h}$  on model levels (i.e., layer boundaries), we must relate the column of any gas to its dry mixing ratio profile, and the pressure interval in each layer must be conceptually divided by assigning fractions of it to the two adjacent levels in such a way that integrating over all levels conserves both total pressure and CO<sub>2</sub> column.

The column of any gas  $X$  may be shown to be given by:

$$C_X = \sum_i \frac{u_i^X}{g_i M_{dry} (1 + \epsilon u_i^{H_2O})} \Delta p_i \quad (3-69)$$

where  $u_i$  is the mixing ratio on level  $i$ ,  $\Delta p_i$  is the pressure interval assigned to that level,  $g$  is the local acceleration due to gravity, and  $M_{dry}$  is the mean molecular weight of dry air (taken to be 0.0289644 kg/mol). It is straightforward to show that this is equivalent to:

$$C_X = \sum_i \frac{u_i^X (1 - q_i)}{g_i M_{dry}} \Delta p_i \quad (3-70)$$

where  $q_i$  is the specific humidity of dry air at level  $i$ . Appendix A of O'Dell et al. [2012] details the calculation of the pressure weighting function on atmospheric levels and is not repeated here.

### 3.6.2 $X_{CO_2}$

Finally,  $X_{CO_2}$  and its uncertainty  $\sigma X_{CO_2}$  are calculated. The  $CO_2$  dry air mole fraction  $X_{CO_2}$  is obtained by averaging the  $CO_2$  profile, weighted by the pressure weighting function  $\mathbf{h}$  such that  $X_{CO_2} = \mathbf{h}^T \hat{\mathbf{x}}$ . The formal error variance in the retrieved  $X_{CO_2}$  is therefore given by:

$$\sigma^2_{X_{CO_2}} = \mathbf{h}^T \hat{\mathbf{S}} \mathbf{h} \quad (3-71)$$

A series of standard error analysis calculations are performed as part of each retrieval to characterize the retrieval results and quantify their uncertainties. The retrieval uncertainties and averaging kernels are calculated from the measurement Jacobian  $\mathbf{K}$  of Section 3.5.1 evaluated at the retrieved state, the measurement covariance matrix, and the *a priori* covariance matrix. From these we compute the column averaging kernel  $\mathbf{a}_{CO_2}$ , the uncertainty due to smoothing and interference  $\tilde{\mathbf{a}}_c$ , the correlation of non- $CO_2$  state vector elements to  $X_{CO_2}$   $\tilde{\mathbf{s}}_1$ , and the components of  $X_{CO_2}$  variance due to measurement error, smoothing, and interference. These last six quantities are defined and discussed below. A complete list of operational retrieval products is given in Table 3-8.

**Table 3-8.** Inverse method products recorded with each sounding.

Name	Size	Comment
$\chi^2$	3	Sum of squares of normalized residuals in each spectrometer
$\hat{\mathbf{x}}$	$N$	Retrieved state vector
$\hat{S}_{ii}$	$n$	Diagonal elements of $\hat{S}$ (error covariance matrix)
$\hat{S}_{CO_2}$	$q^2$	$CO_2$ -only submatrix of $\hat{S}$
$\tilde{\mathbf{s}}_1$	$n-q$	Correlation of $X_{CO_2}$ with non- $CO_2$ elements of $\mathbf{x}$
$\mathbf{A}_{CO_2}$	$q^2$	$CO_2$ -only submatrix of averaging kernel ( $\mathbf{A}$ )
$\mathbf{a}_{CO_2}$	$n$	Column averaging kernel
$\tilde{\mathbf{a}}_c$	$n$	Error in $X_{CO_2}$ due to smoothing and interference
$X_{CO_2}$	1	Column-weighted $CO_2$ dry air mole fraction
$\sigma_m^2$	1	Variance of $X_{CO_2}$ due to measurement noise
$\sigma_s^2$	1	Variance of $X_{CO_2}$ due to smoothing
$\sigma_i^2$	1	Variance of $X_{CO_2}$ due to interference
$\sigma^2 X_{CO_2}$	1	Total variance of $X_{CO_2}$ ( $\sigma_m^2 + \sigma_s^2 + \sigma_i^2$ )
$d_f$	1	Degrees of freedom (full state vector)
$d_{CO_2}$	1	Degrees of freedom ( $CO_2$ profile only)

### 3.6.3 The Column Averaging Kernel $\mathbf{a}_{CO_2}$

Let  $\hat{\mathbf{u}}$  be the retrieved  $CO_2$  mixing ratio profile and  $\hat{\mathbf{e}}$  be the retrieved vector of all non- $CO_2$  quantities:

$$\hat{\mathbf{x}} = \begin{pmatrix} \hat{\mathbf{u}} \\ \hat{\mathbf{e}} \end{pmatrix} \quad (3-72)$$

and let:

$$\mathbf{A}_{ij} = \frac{\partial \hat{\mathbf{x}}_i}{\partial \mathbf{x}_j} \quad i, j = 1, n \quad (3-73)$$

so:

$$\frac{\partial X_{CO_2}}{\partial \mathbf{x}_j} = (\mathbf{h}^T \mathbf{A})_j \quad (3-74)$$

Now consider only the CO<sub>2</sub> part of the state vector, the first  $q$  elements  $u_i$ ,  $i=1$  to  $q$ , and define the column averaging kernel:

$$(\mathbf{a}_{CO_2})_j = \frac{\partial X_{CO_2}}{\partial \mathbf{u}_j} \frac{1}{\mathbf{h}_j} = (\mathbf{h}^T \mathbf{A})_j \frac{1}{\mathbf{h}_j} \quad (3-75)$$

The column averaging kernel  $\mathbf{a}_{CO_2}$  has the property that its elements all equal 1 in the ideal case, where the retrieved  $X_{CO_2}$  responds to changes in  $\mathbf{u}$  exactly as the true value of the profile-weighted mixing ratio. For a real retrieval, the elements of  $\mathbf{a}_{CO_2}$  may be more or less than 1, and will have values much less than 1 in regions where the *a priori* CO<sub>2</sub> profile is important.

### 3.6.4 Smoothing and Interference due to the State Vector $\tilde{\mathbf{a}}_c$

The vector  $\tilde{\mathbf{a}}_c$  captures the smoothing and interference (or cross talk) errors in  $X_{CO_2}$  due to each element of  $\hat{\mathbf{x}}$ . Thus,  $\tilde{\mathbf{a}}_c$  reveals the sensitivity of the retrieved  $X_{CO_2}$  value to uncertainties in elements of the state vector. It may be derived from the full averaging kernel matrix as follows.

The error in  $X_{CO_2}$  is given by:

$$\Delta X_{CO_2} = \mathbf{h}^T (\hat{\mathbf{x}} - \mathbf{x}) = \mathbf{h}^T (\mathbf{A}_{uu} - \mathbf{I})(\mathbf{u} - \mathbf{u}_a) + \mathbf{h}^T \mathbf{A}_{ue}(\mathbf{e} - \mathbf{e}_a) + \varepsilon_u \quad (3-76)$$

Equation (3-76) is an adaptation of Equation (7) in Rodgers and Connor [2003]. The first term in Equation (3-76) represents smoothing error, the second interference error, and  $\varepsilon_u$  all other sources of error. Here,  $\mathbf{A}_{uu}$  and  $\mathbf{A}_{ue}$  are submatrices of  $\mathbf{A}$ , representing the CO<sub>2</sub>-only component and the cross talk components (those which mix elements of the CO<sub>2</sub> profile  $\mathbf{u}$  and the non-CO<sub>2</sub> elements  $\mathbf{e}$ ), respectively.  $\mathbf{I}$  is the identity matrix.

It follows that the error in  $X_{CO_2}$  due to each state vector element is given by:

$$\tilde{\mathbf{a}}_{c,j} = \begin{cases} \mathbf{h}^T (\mathbf{A} - \mathbf{I}) \sigma_j, & j = 1, q \\ \mathbf{h}^T \mathbf{A} \sigma_j, & j = q+1, n \end{cases} \quad (3-77)$$

where  $\sigma_j$  is the error in element  $j$ . Since  $h_j \equiv 0$  for  $j > q$ , the full matrix  $\mathbf{A}$  may be used in place of the submatrices in Eq. (3-76).

Alternatively, Eq. (3-77) may be written:

$$\tilde{\mathbf{a}}_{c,j} = \begin{cases} \left( \frac{\partial X_{CO_2}}{\partial \mathbf{u}_j} - \mathbf{h}_j \right) \sigma_j = (\mathbf{a}_{CO_2,j} - 1) \mathbf{h}_j \sigma_j, j = 1, q \\ \frac{\partial X_{CO_2}}{\partial \mathbf{e}_j} \sigma_j, j = q+1, n \end{cases} \quad (3-78)$$

The first  $q$  elements of  $\tilde{\mathbf{a}}_c$ , corresponding to the CO<sub>2</sub> profile, are components of the smoothing error. The remaining elements represent the interference, or cross talk error.

### 3.6.5 Correlation of X<sub>CO2</sub> with non-CO<sub>2</sub> State Vector Elements $\tilde{\mathbf{s}}_1$

We define  $\tilde{\mathbf{s}}_1$ , the correlation of X<sub>CO2</sub> with the state vector elements, to aid the diagnosis and understanding of cross talk. Define a matrix  $\mathbf{H}$  with dimension  $n \times (n+1)$  such that:

$$\mathbf{H}_{ij} = \begin{cases} \mathbf{h}_i, j = 1 \\ 1, i = j - 1, j > 1 \\ 0, else \end{cases} \quad (3-79)$$

Thus:

$$\mathbf{H}^T \hat{\mathbf{x}} = \begin{pmatrix} X_{CO_2} \\ x_1 \\ x_2 \\ \cdot \\ \cdot \\ x_n \end{pmatrix} \quad (3-80)$$

and the correlation matrix of  $\mathbf{H}^T \hat{\mathbf{x}} \hat{\mathbf{x}}$  is given by  $\mathbf{H}^T \hat{\mathbf{C}} \mathbf{H}$ , where  $\hat{\mathbf{C}}$  is the correlation matrix corresponding to  $\hat{\mathbf{S}}$ .

$\tilde{\mathbf{s}}_1$  is the first row of  $\mathbf{H}^T \hat{\mathbf{C}} \mathbf{H}$ , where  $\tilde{\mathbf{s}}_1 = (1 \text{ pX}_{CO_2,j} \dots)$  for  $j = 1, n$ . The second and subsequent elements of  $\tilde{\mathbf{s}}_1$  are the correlation coefficients of X<sub>CO2</sub> with each element of the state vector.

### 3.6.6 Components of X<sub>CO2</sub> Variance

In Section 3.6.2 above, the variance in X<sub>CO2</sub> is given in Equation (3-71) for pressure weighting function  $\mathbf{h}$  and *a posteriori* error covariance  $\hat{\mathbf{S}}$ . We also calculate the component parts of the variance, as described next.

In Equation (3.16) of Rodgers [2000], the total error is broken up into measurement, smoothing, and forward model error (which includes both radiative transfer and the instrument). The operational L2 code has no explicit information on forward model error, though it may be implicitly included in the measurement error estimate provided as input. Formally then, we separate the error into measurement and smoothing components. Further, smoothing due to the full state vector may logically be separated into smoothing due to CO<sub>2</sub> itself and interference due to non-CO<sub>2</sub> elements, as described in Rodgers and Connor [2003]. Thus, we calculate variance components due to spectral error (measurement and implicit model error), smoothing (due to CO<sub>2</sub>), and interference.

First, define the contribution function as:

$$\mathbf{G}_y = \hat{\mathbf{S}}\mathbf{K}^T\mathbf{S}_\varepsilon^{-1} \quad (3-81)$$

then the measurement error covariance is:

$$\hat{\mathbf{S}}_m = \mathbf{G}_y\mathbf{S}_\varepsilon\mathbf{G}_y^T \quad (3-82)$$

the smoothing error due to CO<sub>2</sub> is:

$$\hat{\mathbf{S}}_s = (\mathbf{A}_{\text{CO}_2} - \mathbf{I})\mathbf{S}_{a,\text{CO}_2}(\mathbf{A}_{\text{CO}_2} - \mathbf{I})^T \quad (3-83)$$

and the interference error is:

$$\hat{\mathbf{S}}_i = \mathbf{A}_{ue}\mathbf{S}_{ae}\mathbf{A}_{ue}^T \quad (3-84)$$

where subsets of the averaging kernel  $\mathbf{A}$  and *a priori* covariance  $\mathbf{S}_a$  are defined as follows:

$$\mathbf{A}_{\text{CO}_2} = \mathbf{A}(\mathbf{1:nz}, \mathbf{1:nz}) \text{ for } \underline{\text{nz}} = \# \text{ levels in CO}_2 \quad (3-85)$$

$$\underline{\mathbf{A}}_{ue} = \mathbf{A}(\mathbf{1:nz}, \underline{\text{nz}}+1:\mathbf{n}) \quad \text{for } \mathbf{n} = \text{full length of state vector} \quad (3-86)$$

$$\mathbf{S}_{a\text{CO}_2} = \mathbf{S}_a(\mathbf{1:nz}, \mathbf{1:nz}) \quad (3-87)$$

$$\underline{\mathbf{S}}_{ae} = \underline{\mathbf{S}}_a(\underline{\text{nz}}+1:\mathbf{n}, \underline{\text{nz}}+1:\mathbf{n}) \quad (3-88)$$

Finally, the variance components are determined. The measurement error is given by:

$$\sigma_m^2 = \underline{\mathbf{h}}^T \underline{\hat{\mathbf{S}}}_m \underline{\mathbf{h}} \quad (3-89)$$

the smoothing due to CO<sub>2</sub> is given by:

$$\sigma_s^2 = \underline{\mathbf{h}}^T \underline{\hat{\mathbf{S}}}_s \underline{\mathbf{h}} \quad (3-90)$$

and the interference is given by:

$$\sigma_i^2 = \underline{\mathbf{h}}^T \underline{\hat{\mathbf{S}}}_i \underline{\mathbf{h}} \quad (3-91)$$

These components of variance have the property that:

$$\sigma_{X\text{CO}_2}^2 = \sigma_m^2 + \sigma_s^2 + \sigma_i^2 \quad (3-92)$$



## 4 Algorithm Specification

The purpose of this section is to describe implementation details of the remote sensing algorithms that produce the OCO-2 Level 2 end-user Diagnostic (L2Dia) and Standard (L2Std) data products. The science basis for these algorithms is provided in Section 3 of this document. The algorithm specification provides a software description of that science as implemented in the operational ground system—the Science Data Operations System (SDOS). The intent of the algorithm specification is to capture the as-built operational implementation of the algorithm.

### 4.1 Data System Context

The SDOS Processing System consists of a set of Product Generation Executables (PGEs) and the infrastructure that initiates and routes data between them. Conceptually, the PGEs can be grouped into processing levels.

- Level 0 Processing prepares incoming datasets for higher-level processing.
- Level 1 Processing generates engineering data products and calibrated, geolocated science measurements.
- Level 2 Preprocessing and Sounding Selection selects the soundings for which to analyze in terms of  $X_{CO_2}$ . SIF retrievals are also generated in the preprocessing step.
- Level 2 Processing generates the  $X_{CO_2}$  science results.

A full description of the L2 PGEs is provided in Section 4.7 below.

### 4.2 Level 2 Preprocessing

As stated above, three preprocessing steps are run on the OCO-2 data, but within L2. The first of these is the ABO2 Preprocessor whose main function is to identify heavily cloudy and aerosol-laden scenes for exclusion from FP analysis, but plays other roles as well. The IMAP preprocessor serves as a complementary cloud-screening algorithm, and simultaneously makes retrievals of SIF. The Sounding Selection algorithm takes input from both of these processors, as well as L1B data and meteorological data in order to select which soundings to run through the FP algorithm. This is because the FP algorithm is computationally intensive, and it is not currently feasible to run all soundings through it, so care must be taken to only select soundings with a high likelihood of yielding useful  $X_{CO_2}$  retrievals.

#### 4.2.1 ABO2 Preprocessor

The  $O_2$  A band cloud screening algorithm was developed at Colorado State University under the ACOS program. It employs a fast Bayesian retrieval to estimate surface pressure and surface albedo, assuming clear-sky conditions with only molecular Rayleigh scattering, from high-resolution spectra of the  $O_2$  A band near 765 nm. The estimated surface pressure, surface albedo and the chi-squared goodness-of-fit statistic are used to flag scenes as cloudy, clear, or indeterminate [Taylor et al., 2011]. Further details of the ABO2 technique are described in the ABO2 Algorithm ATBD as well as an upcoming publication [Taylor et al., 2015, *in prep*].

The basic method is that, absent clouds or aerosols, the surface pressure of a clear scene can be determined to within 2 – 5 hPa accuracy using the  $O_2$  A band spectrum of reflected sunlight. This is because of the strong oxygen absorption features that are present in this band. When surface pressure is higher, the absorption features are deeper for a given observation geometry.

When clouds or aerosols are present, they change the path lengths for most photons, either via shortening or lengthening, such that the retrieved surface pressure can be very different from the expected value based on a meteorological forecast, and also the O<sub>2</sub> A spectrum itself cannot be well-fit with a clear-sky assumption.

The key screening parameters are the retrieved minus *a priori* surface pressure, as well as the goodness of fit as indicated through a  $\chi^2$  metric. Over ocean away from glint observations, the retrieved surface albedo is also used (because oceans are only bright due to the presence of clouds or aerosols away from the glint spot). These parameters interpreted internally via thresholds to provide a simple cloud flag. As the primary objective of the ABO2 preprocessor is to flag as cloudy scenes that are obviously cloudy or aerosol-contaminated, thresholds are set rather loosely so scenes that have some cloud or aerosol contamination will sometimes pass the filter. Furthermore, the key screening parameters are used as inputs to the genetic algorithm used for sounding selection [Mandrake, 2013], described in section 4.2.2.

#### 4.2.2 IMAP Preprocessor (and SIF Retrievals)

The IMAP-DOAS pre-processor is a fast non-scattering retrieval algorithm that fits vertical column densities of CO<sub>2</sub> and H<sub>2</sub>O [Frankenberg et al., 2005] as well as a Fraunhofer-line based SIF retrieval algorithm based on Frankenberg et al. [2011]. CO<sub>2</sub> and H<sub>2</sub>O fits are performed independently in the weak and strong CO<sub>2</sub> bands, respectively. More details can be found in the IMAP Algorithm ATBD.

The key screening parameters are both the CO<sub>2</sub> and well as the H<sub>2</sub>O ratios, which are the ratios of retrieved CO<sub>2</sub> and H<sub>2</sub>O columns in the weak and strong band. These ratios deviate from unity if significant scattering is present in the atmosphere and can thus be used to identify scattering on top of the ABO2 cloud-screen and both have been shown to be complementary.

The SIF fits are performed independently at 757 and 771 nm and will be provided in the IMAP-DOAS L2 files.

### 4.3 Level 2 Full Physics Inputs

The L2 software utilizes both the L1B and GEOS FP-IT resampled products generated earlier in the processing pipeline. The L1B file provides radiometrically corrected spectra and geolocation information. The GEOS FP-IT Resampled product provides pressure, temperature and specific humidity profiles derived from GEOS FP-IT forecasted or reinterpreted products as described in Section 3.3.2.2. The GEOS FP-IT data are interpolated in time and location along the OCO-2 orbit track to match the L1B geolocation. More information about these products can be found in the L1B Algorithm Specification Document. The L2 software also utilizes the additional static data sources:

- ABSCO tables—tabulated absorption coefficients providing gas optical properties information, as described in Section 3.2.1
- Solar spectrum information, as described in Section 3.2.2
- *A priori* CO<sub>2</sub> profiles, as described in Section 3.3.2.1
- MERRA *a priori*—aerosol optical depth data for the world interpreted on a monthly bases from MERRA products, as described in Section 3.3.2.3
- Aerosol Properties—scattering and optical properties of scattering particles, derived from MERRA data, as described in Section 3.3.2.3
- *a priori* surface properties, as described in Section 0
- *a priori* wavelength grid offset, as described in Section 3.3.2.5

- Residual empirical orthogonal function parameters, as described in Section 3.3.2.6
- *a priori* values for atmospheric components not coming from other products
  - Fluorescence
  - Coxmunk windspeed, Lambertian albedo
  - Zero Level Offset
- *a priori* covariance values for all quantities, as described in Section 3.3.2

#### 4.3.1 Requirements on Inputs

The L1B and GEOS FP-IT input data products are required to be:

- Time-ordered—soundings occur in increasing time order
- HDF5—Hierarchical data format v5
- Single operating mode—only nadir, glint, or target soundings in one file
- Single orbit—only a single orbit of data
- Quality control flags set—quality flags do not have fill values
- Geolocated—geolocation has already been performed on pointing information and is included in L1B files

#### 4.3.2 Attributes of Input Products

Attributes of the L1B Science and GEOS FP-IT input data products are given in Table 4-1 and Table 4-2, respectively.

**Table 4-1.** L1B Science product attributes.

L1B Science Product	
Description	The spectra in the L1B product include factors that (1) correct for the curvature of the signal on the instrument image plane, (2) apply the radiometric calibration to the detector counts in the science telemetry, and (3) eliminate the effects of spatial cross talk in the spectrometer signal
Long Name	Spatially ordered, geolocated, calibrated science spectra
Frequency	1 per contiguous observation mode in a single orbit (sample resolution science data only)
Format	HDF5
Average orbit volume	1.37 GB

**Table 4-2.** L2Met product attributes.

L2Met Product	
Description	GEOS FP-IT forecast data resampled to the location of the OCO-2 measurements
Long Name	Spatially ordered, geolocated meteorological forecasts
Frequency	1 per contiguous observation mode in a single orbit (sample resolution science data only)
Format	HDF5
Average orbit volume	1.60 GB

#### 4.4 Level 2 Full Physics Outputs

The L2 FP PGE produces a single L2 FP output file in HDF format for each sounding. These files are then aggregated by the L2Std PGE into products destined for distribution at the Distributed Active Archive Center.

Attributes of the L2 Single Sounding, L2Dia, and L2Std output data products are given in Table 4-3, Table 4-4, and Table 4-5, respectively.

**Table 4-3. L2 Single Sounding product attributes.**

<b>L2 Single Sounding Product</b>	
Description	Direct output of L2 FP process per sounding. Contains $X_{CO_2}$ , atmospheric profiles, error estimates, and the state vector
Long Name	L2 FP single sounding output
Frequency	1 per sounding processed
Format	HDF5
Average orbit volume	1.0 GB
Data groups*	Product Metadata RetrievalHeader RetrievalResults RetrievedStateVector SpectralParameters

\*see Data User's Guide

**Table 4-4. L2Dia product attributes.**

<b>L2 Diagnostic Product</b>	
Description	Geolocated retrieved state vectors with $CO_2$ column averaged dry air mole fraction - physical model. Geolocated retrieved $CO_2$ column averaged dry air mole fraction - physical model.
Long Name	Geolocated $X_{CO_2}$ and state vectors - physical model
Frequency	1 per contiguous observation mode in a single orbit (sample resolution science data only)
Format	HDF5
Average orbit volume	2.09 GB
Data groups*	Product Metadata L1bScSoundingReference RetrievalHeader RetrievalGeometry PreprocessingResults RetrievalResults AlbedoResults DispersionResults AerosolResults RetrievedStateVector SpectralParameters

\*see Data User's Guide

**Table 4-5.** L2Std product attributes.

<b>L2 Standard Product</b>	
Description	Geolocated retrieved CO <sub>2</sub> column averaged dry air mole fraction - physical model
Long Name	Geolocated X <sub>CO2</sub> - physical model
Frequency	1 per contiguous observation mode in a single orbit (sample resolution science data only)
Format	HDF5
Average orbit volume	1.32 GB
Data groups*	Product Metadata L1bScSoundingReference RetrievalHeader RetrievalGeometry PreprocessingResults RetrievalResults AlbedoResults DispersionResults AerosolResults SpectralParameters

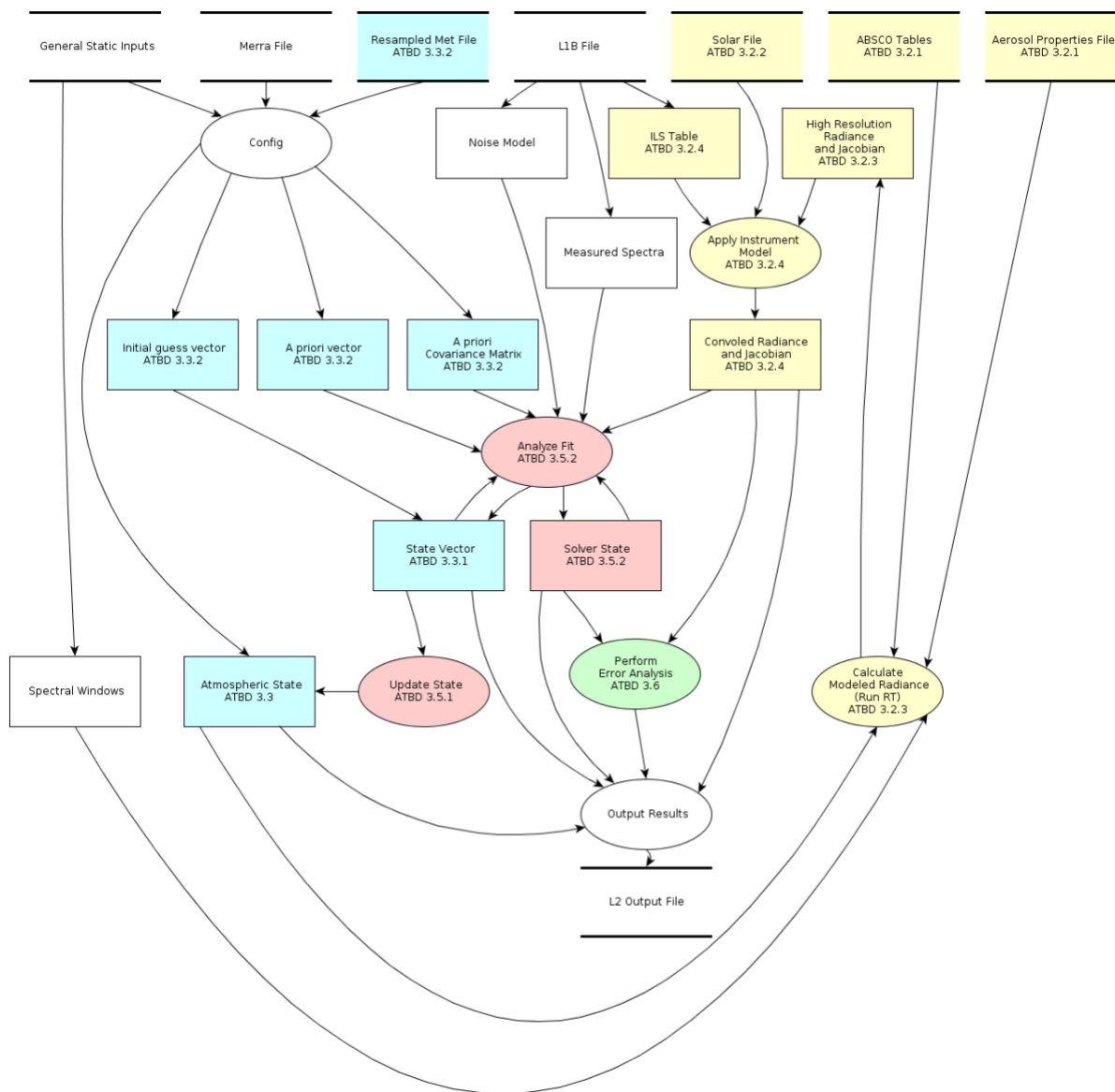
\*see Data User's Guide

#### 4.5 Description of Full Physics Major Code Sections

The FP code carries out the following steps to process a single sounding:

- Load Lua configuration, as described in Section 4.6
- Set up inputs as described in Section 4.3.2 and class connections from configuration
- Perform radiative transfer, as described in Section 3.2.3
- Apply solar model on spectra, as described in Section 3.2.2
- Model instrument effects on spectra, as described in Section 3.2.4
- Non-linear least squares inversion to compare modeled and measured spectra, as described in Section 3.5
- Update state vector, as described in Section 3.5.1
- Push state vector updates into models, as described in Section 3.3
- Check for convergence, as described in Section 3.5.2
- If converged perform error analysis, as described in Section 3.6
- Write L2 Single Sounding Product output file, as described in Section 4.4

Figure 4-1 shows the flow of data through the various L2 FP software components. It is meant to give a more detailed picture of how the L2 code is organized. A diagram of the algorithm control flow can be seen in Figure 4-2.



**Figure 4-1.** Data flow through the L2 FP software components.

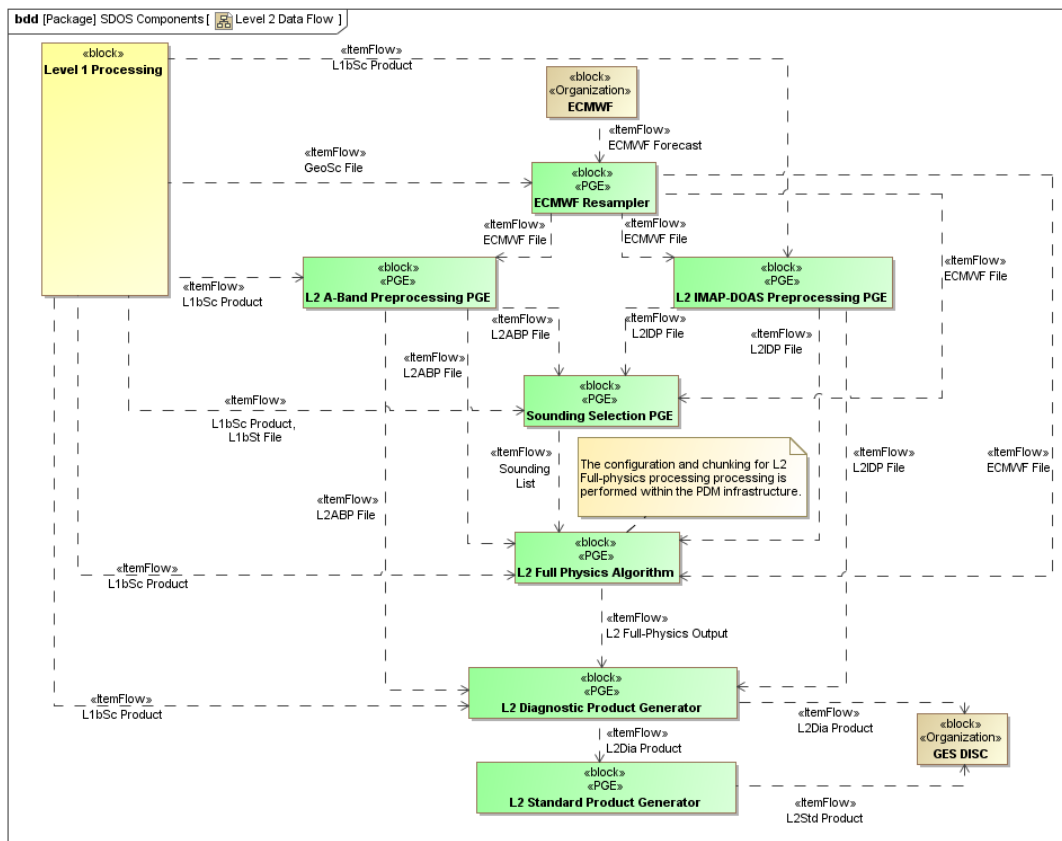
## 4.6 Lua Configuration

Lua is a lightweight, embeddable scripting language. It was chosen to manage the L2 FP configuration over a declarative approach due to its flexibility. The L2 code itself was written in a manner such that the C++ portions are components that have minimal dependency on each other and adhere to well defined interfaces. The Lua configuration selects which components are used and makes them aware of each other. The configuration files themselves are written in layers where more specific configurations extend the general. At the bottom level are configurations that implement instrument specific arrangements of L2 FP components. These files specify which files are involved in the processing as well as algorithmic choices. At this level the Lua code looks like declarative keyword/value configuration files. This approach combines flexibility with simplicity for end users.

The configuration loads input data, building *a priori* and initial guess vectors as it processes components. Components are connected to each other and passed their input in a hierarchal manner. Dependent components are loaded first so they are initialized and can be passed to components that need them. The top most level is the OCO forward model component. It is created last and when ready called to start iterative execution as described in previous sections.

## 4.7 Level 2 Processing PGEs

A schematic illustrating the L2 Processing PGEs is given in Figure 4-2 below.



**Figure 4-2. L2 Processing overview.**

### 4.7.1 GEOS FP-IT Resampler

Purpose: The GEOS FP-IT Resampler extracts the meteorological data relevant to a single orbit from a global GEOS FP-IT dataset.

### Connectivity:

- Receives a GEOS FP-IT Forecast file from GEOS FP-IT
- Receives a GeoSc file from the Geolocation PGE
- Provides a GEOS FP-IT file to the L2 A-Band Preprocessing PGE, the L2 Iterative Maximum *a Posteriori* Differential Optical Absorption Spectroscopy (IMAP-DOAS) Preprocessing PGE, the Sounding Selection PGE, and the L2 FP algorithm

Multiplicity: The Processing System runs the GEOS FP-IT Resampler once for each L1BSc product.



#### 4.7.2 L2 A-Band Preprocessing PGE

Purpose: The L2 A-Band Preprocessing PGE derives attributes of each sounding using a retrieval on the ABO2 measurement only.

Connectivity:

- Receives an L1BSc product from the CalApp PGE
- Receives an GEOS FP-IT file from the GEOS FP-IT Resampler
- Provides an L2ABP file to the Sounding Selection PGE, the L2 FP algorithm, and the L2Dia Product Generator

Multiplicity: The Processing System runs the L2 A-Band Preprocessing PGE once for each L1BSc product.

#### 4.7.3 L2 IMAP-DOAS Preprocessing PGE

Purpose: The L2 IMAP-DOAS Preprocessing PGE derives attributes of each sounding using an IMAP-DOAS retrieval.

Connectivity:

- Receives an L1BSc product from the CalApp PGE
- Receives an GEOS FP-IT file from the GEOS FP-IT Resampler
- Provides an L2IDP file to the Sounding Selection PGE, the L2 FP algorithm, and the L2 Diagnostic Product Generator

Multiplicity: The Processing System runs the L2 IMAP-DOAS Preprocessing PGE once for each L1BSc product.

#### 4.7.4 Sounding Selection PGE

Purpose: The Sounding Selection PGE identifies measurements for the L2 FP Algorithm to use to generate X<sub>CO2</sub> values.

Connectivity:

- Receives an L1BSc product from the CalApp PGE
- Receives an L1BSt file from the L1BStat PGE
- Receives an GEOS FP-IT file from the GEOS FP-IT Resampler
- Receives an L2ABP file from the L2 A-Band Preprocessing PGE
- Receives an L2IDP file from the L2 IMAP-DOAS Preprocessing PGE
- Provides an L2Lst file to the L2 FP algorithm
- Generates an L2Sel file that is used for analysis only

Multiplicity: The Processing System runs the Sounding Selection PGE once for each L1BSc product.

#### 4.7.5 L2 FP Algorithm

Purpose: The L2 FP algorithm retrieves X<sub>CO2</sub> for selected science measurements.

Connectivity:

- Receives an L1BSc product from the CalApp PGE
- Receives an GEOS FP-IT file from the GEOS FP-IT Resampler
- Receives an FP Configuration file from the FP Populator

- Receives an L2ABP file from the L2 A-Band Preprocessor PGE
- Receives an L2IDP file from the L2 IMAP-DOAS Preprocessor PGE
- Generates an L2 FP Output file for use by the FP Diagnostic Product Generator

**Multiplicity:** The Processing System runs the L2 FP algorithm once for each measurement to be processed.

#### 4.7.6 FP Diagnostic Product Generator

**Purpose:** The FP Diagnostic Product Generator extracts information from The L1BSc product, the L2ABP file, the L2IDP file, and all the L2 FP Output files for an orbit and saves that information to a single file.

**Connectivity:**

- Receives an L1BSc product from the CalApp PGE
- Receives an L2ABP file from the L2 A-Band Preprocessor PGE
- Receives an L2IDP file from the L2 IMAP-DOAS Preprocessor PGE
- Receives a set of L2 FP Output files from the L2 FP Algorithm, one file for every measurement processed for a single orbit
- Produces an L2Dia product for use by the FP Standard Product Generator and delivery to the GES DISC

**Multiplicity:** The Processing System runs the FP Diagnostic Product Generator once for each L1BSc product.

#### 4.7.7 FP Standard Product Generator

**Purpose:** The FP Standard Product Generator subsets the information contained in an L2Dia product to generate an L2Std product.

**Connectivity:**

- Receives an L2Dia product from the FP Diagnostic Product Generator
- Produces an L2Std product for delivery to the GES DISC

**Multiplicity:** The Processing System runs the FP Standard Product Generator once for each L1BSc product.

#### 4.7.8 Level 2 PGE Multiplicity Summary

Table 4-6 summarizes the number of instances of L2 PGEs run during a nominal day.

**Table 4-6.** Daily L2 PGE instances.

PGE	Total
GEOS FP-IT Resampler	1/L1BSc product
L2 A-Band Preprocessor PGE	1/L1BSc product
L2 IMAP-DOAS Preprocessor PGE	1/L1BSc product
Sounding Selection PGE	1/L1BSc product
FP Populator	1/chunk
L2 FP algorithm	1/measurement
FP Diagnostic Product Generator	1/L1BSc product
FP Standard Product Generator	1/L1BSc product

## 4.8 Error Handling

Inputs for ranges and expected units checked. For example, ensuring that latitude and longitude values are within acceptable ranges.

Radiative transfer checks sanity of inputs. For example, checking that albedo is less than 1.0 or that optical depths are at acceptable levels.

Math overflow and underflow errors caught. These are caught as floating point exceptions. Additionally there are checks to flag when NaN values are created.

Exception handling catches unexpected errors using mechanisms built into C++. Many exceptions are caught within individual components and dealt with if their possible occurrences are known ahead of time. However, third party code might throw unexpected exceptions that are rare. These would be propagated to the top level of the software. The code will try to write any output data pending before existing. It also marks the output product with the *.error* extension.

## 4.9 Commercial off-the-Shelf Components

The Level 2 FP implementation uses a number of third party software packages. These are as follows:

- Blitz++ — <http://blitz.sourceforge.net/>
- Boost C++ Libraries — <http://www.boost.org/>
- Bzip2 — <http://www.bzip.org/>
- GNU Scientific Library — <https://www.gnu.org/software/gsl/>
- HDF5 — <http://www.hdfgroup.org/HDF5/>
- Lua — <http://www.lua.org/>
- Luabind — <http://www.rasterbar.com/products/luabind.html>
- Nose Python Unit Testing — <https://nose.readthedocs.org/en/latest/>

## 4.10 Quality Assessment and Recording

There are no quality assessments made in the FP Single Sounding files themselves. Instead the L2Std Product PGE assigns a quality assessment, referred to as a Warn Level, using data from the L2Dia Product. For more details on the Warn Level generation procedure, see Mandrake et al. [2012].

## 5 References

- Allen, C.W. (1964), *Astrophysical Quantities*, Athlone Press (2<sup>nd</sup> ed.), London.
- Anderson, P. W. (1949), Pressure Broadening in the Microwave and Infra-Red Regions, *Phys. Rev.*, vol. 76, Issue 5, pp. 647-661.
- Andrews, A., Boering, K., Daube, B., Wofsy, S., Loewenstein, M., Jost, H., Podolske, J., Webster, C., Herman, R., Scott, D., Flesch, G. J., Moyer, E. J., Elkins, J. W., Dutton, G. S., Hurst, D. F., Moore, F. L., Ray, E. A., Romashkin, P. A., and Strahan, S. E. (2001). Mean ages of stratospheric air derived from in situ observations of CO<sub>2</sub>, CH<sub>4</sub>, and N<sub>2</sub>O, *J. Geophys. Res.-Atmos.*, 106, 32295–32314.
- Barton, John J., and Lee R. Nackman (1997), *Scientific and Engineering C++: An Introduction with Advances Techniques and Examples*. Addison-Wesley.
- Baum, B. A., A. J. Heymsfield, P. Yang, and S. T. Bedka (2005a), Bulk scattering properties for the remote sensing of ice clouds. Part I: Microphysical data and models, *J. Appl. Meteorol.*, 44, 1885–1895.
- Baum, B. A., P. Yang, A. J. Heymsfield, S. Platnick, M. D. King, Y.-X. Hu, and S. T. Bedka (2005b), Bulk scattering properties for the remote sensing of ice clouds. Part II: Narrowband models, *J. Appl. Meteorol.*, 44, 1896–1911.
- Baum, B. A., P. Yang, A. J. Heymsfield, A. Bansemer, B. H. Cole, A. Merrelli, C. Schmitt, and C. Wang (2014), Ice cloud single-scattering property models with the full phase matrix at wavelengths from 0.2 to 100 $\mu$ m. *J. Quant. Spectrosc. Radiat. Transfer*, in press.
- Benner, D. C., V. M. Devi, E. Nugent, K. Sung, L. R. Brown, C. E. Miller and R. A. Toth (2011), Line parameters of Carbon Dioxide in the 4850 cm<sup>-1</sup> region, poster presentation at the 22<sup>nd</sup> Colloquium on High Resolution Molecular Spectroscopy, Dijon, France.
- Benner, D. C., Devi, V. M., Sung, K., Brown, L. R., Miller, C. E., Payne, V. H., Drouin, B. J., Yu, S., Crawford, T. J., Mantz, A. W., Smith, M. A. H., and Gamache, R. R. (2016), Line parameters including temperature dependences of air- and self-broadened line shapes of CO<sub>2</sub>: 2.06- $\mu$ m region, *Journal of Molecular Spectroscopy*, 326, 21–47, doi:10.1016/j.jms.2016.02.012.
- Boesch, H., G.C. Toon, B. Sen, R.A. Washenfelder, P.O. Wennberg, M. Buchwitz, R. de Beek, J.P. Burrows, D. Crisp, M. Christi, B.J. Connor, V. Natraj, and Y.L. Yung (2006), Space-based near-infrared CO<sub>2</sub> measurements: Testing the Orbiting Carbon Observatory retrieval algorithm and validation concept using SCIAMACHY observations over Park Falls, Wisconsin, *Journal of Geophysical Research-Atmospheres*, 111 (D23).
- Brown, L. R., and C. Plymate (2000), Experimental line parameters of the Oxygen A band at 760 nm, *J. Mol. Spectrosc.* 100, 166-179.
- Brown, L. (2009), Line list is available in the acos/absco repository at the following location: acos/level\_2/exe/absco/input/line\_mixing/o2/lm\_o2\_inputs\_20091026/SDF.dat

- Butz, A., S. Guerlet, O. Hasekamp, D. Schepers, A. Galli, I. Aben, C. Frankenberg et al. (2011), Toward accurate CO<sub>2</sub> and CH<sub>4</sub> observations from GOSAT. *Geophysical Research Letters* 38, no. 14.
- Chandrasekhar, S., (1950), Radiative Transfer. Oxford Univ. Press, Oxford.
- Chandrasekhar, S., (1960), Radiative Transfer. Dover Publications, Dover, N.Y.
- Connor B. J., H. Boesch, G. Toon, B. Sen, C. Miller and D. Crisp (2008), Orbiting Carbon Observatory: Inverse method and prospective error analysis, *J. Geophys. Res.*, 113, D05305, doi: 10.1029/2006JD008336.
- Cox, C. and W. Munk (1954), Measurement of the roughness of the sea surface from photographs of the Sun's glitter, *J. Opt. Soc. Am.*, 44, 838-850.
- Crisp, D., R.M. Atlas, F.-M. Breon, L.R. Brown, J.P. Burrows, P. Ciais, B.J. Connor, S.C. Doney, I.Y. Fung, D.J. Jacob, C.E. Miller, D. O'Brien, S. Pawson, J.T. Randerson, P. Rayner, R.J. Salawitch, S.P. Sander, B. Sen, G.L. Stephens, P.P. Tans, G.C. Toon, P.O. Wennberg, S.C. Wofsy, Y.L. Yung, Z. Kuang, B. Chudasama, G. Sprague, B. Weiss, R. Pollock, D. Kenyon, S. Schroll (2004), The Orbiting Carbon Observatory (OCO) mission, *Advances in Space Research* 34, 700–709.
- Crisp, D., C. E. Miller, and P. L. DeCola (2007), NASA Orbiting Carbon Observatory: Measuring the column averaged carbon dioxide mole fraction from space, *JARS*, 12/1/2007.
- Crisp, D. (2008), The Orbiting Carbon Observatory: NASA's First Dedicated Carbon Dioxide Mission, *Proc. SPIE*, 7106, DOI: 10.1117/12.802194.
- Crisp, D., Fisher, B. M., O'Dell, C., Frankenberg, C., Basilio, R., Boesch, H., Brown, L. R., Castano, R., Connor, B., Deutscher, N. M., Eldering, A., Griffith, D., Gunson, M., Kuze, A., Mandrake, L., McDuffie, J., Messerschmidt, J., Miller, C. E., Morino, I., Natraj, V., Notholt, J., O'Brien, D., Oyafo, F., Polonsky, I., Robinson, J., Salawitch, R., Sherlock, V., Smyth, M., Suto, H., Taylor, T., Thompson, D. R., Wennberg, P. O., Wunch, D., and Yung, Y. L. (2012). The ACOS XCO<sub>2</sub> retrieval algorithm, Part 2: Global XCO<sub>2</sub> data characterization, *Atmos. Meas. Tech.*, 5, 687-707.
- Crisp, David, B. Fisher, Christopher O'Dell, Christian Frankenberg, R. Basilio, H. Bosch, L. R. Brown et al. (2012), The ACOS CO<sub>2</sub> retrieval algorithm-Part II: Global XCO<sub>2</sub> data characterization, 687.
- Crowell, S., Baker, D., Schuh, A., Basu, S., Jacobson, A., Chevallier, F., Liu, J., Deng, F., Feng, L., McKain, K., Chatterjee, A., Miller, J., Stephens, B., Eldering, A., Crisp, D., Schimel, D., Nassar, R., O'Dell, C., Oda, T., Sweeney, C., Palmer, P., and Jones, D., (2019), The 2015-2016 Carbon Cycle as Seen from OCO-2 and the Global In Situ Network, *Atmos. Chem. Phys.*, 19: 7347–7376, 2019, doi:10.5194/acp-19-7347-2019.
- de Rooij, W. A., and C. C. A. H. van der Stap (1984), Expansion of Mie scattering matrices in generalized spherical functions, *Astron. Astrophys.*, 131, 237–248.
- Deutscher, N. M. et al. (2010), Total column CO<sub>2</sub> measurements at Darwin, Australia - Site description and calibration against in situ aircraft profiles, *Atmos. Meas. Tech.*, 3(4), 947–958, doi:10.5194/amt-3-947-2010.

- Devi, V. M., D. C. Benner, L. R. Brown, C. E. Miller, and R. A. Toth (2007a), Line mixing and speed dependence in CO<sub>2</sub> at 6348 cm<sup>-1</sup>: Constrained multispectrum analysis of intensities and line shapes in the 30012 ← 00001 band, *J. Mol. Spectrosc.* 242, 90-117.
- Devi, V. M., D. C. Benner, L. R. Brown, C. E. Miller, and R. A. Toth (2007b), Line mixing and speed dependence in CO<sub>2</sub> at 6227.9 cm<sup>-1</sup>: Constrained multispectrum analysis of intensities and line shapes in the 30013 ← 00001 band, *J. Mol. Spectrosc.* 245, 52-80.
- Devi, V. M., Benner, D. C., Sung, K., Brown, L. R., Crawford, T. J., Miller, C. E., Drouin, B. J., Payne, V. H., Yu, S., Smith, M. A. H., Mantz, A. W., and Gamache, R. R. (2016), Line parameters including temperature dependences of self- and air-broadened line shapes of CO<sub>2</sub>: 1.6-μm region, *Journal of Quantitative Spectroscopy & Radiative Transfer*, 177, 117–144, doi:10.1016/j.jqsrt.2015.12.020.
- Dlugokencky, E.J., Mund, J.W., Crotwell, A.M., Crotwell, M.J., and Thoning, K.W. (2019), Atmospheric Carbon Dioxide Dry Air Mole Fractions from the NOAA ESRL Carbon Cycle Cooperative Global Air Sampling Network, 1968-2018, Version: 2019-07, <https://doi.org/10.15138/wkgj-f215>
- Dubovik, O. et al. Statistically optimized inversion algorithm for enhanced retrieval of aerosol properties from spectral multi-angle polarimetric satellite observations. *Atmos. Meas. Tech.* 4, 975–1018 (2011).
- Dufour, E. and F.-M. Bréon, (2003), Spaceborne Estimate of Atmospheric CO<sub>2</sub> Column by Use of the Differential Absorption Method: Error Analysis, *Appl. Opt.* 42, 3595-3609.
- Drouin, B. J., V. H. Payne, F. Oyafo, K. Sung and E. J. Mlawer (2014), Pressure broadening of oxygen by water, *Journal of Quantitative Spectroscopy and Radiative Transfer*, Volume 133, January 2014, Pages 190-198, ISSN 0022-4073.
- Drouin, B. J., Benner, D. C., Brown, L. R., Cich, M. J., Crawford, T. J., Devi, V. M., Guillaume, A., Hodges, J. T., Mlawer, E. J., Ro-bichaud, D. J., Oyafo, F., Payne, V. H., Sung, K., Wishnow, E. H., and Yu, S. (2017), Multispectrum analysis of the oxygen A-band, *Journal of Quantitative Spectroscopy & Radiative Transfer*, 186, 118–138, doi:10.1016/j.jqsrt.2016.03.037.
- Eldering, A., Taylor, T. E., O'Dell, C. W., and Pavlick, R.: The OCO-3 mission; measurement objectives and expected performance based on one year of simulated data, *Atmos. Meas. Tech. Discuss.*, <https://doi.org/10.5194/amt-2018-357>, in review, 2018.
- Frankenberg, C., Platt, U., and Wagner, T. (2005), Iterative maximum *a posteriori* (IMAP-) DOAS for retrieval of strongly absorbing trace gases: Model studies for CH<sub>4</sub> and CO<sub>2</sub> retrieval from near- infrared spectra of SCIAMACHY onboard ENVISAT, *Atmos. Chem. Phys.*, 5, 9–22.
- Frankenberg, C., Fisher, J., Worden, J., Badgley, G., Saatchi, S., Lee, J.-E., et al. (2011), New global observations of the terrestrial carbon cycle from GOSAT: Patterns of plant fluorescence with gross primary productivity. *Geophysical Research Letters*, 38 (17), L17706.

- Frankenberg, C., et al. (2012), Remote sensing of near-infrared chlorophyll fluorescence from space in scattering atmospheres: implications for its retrieval and interferences with atmospheric CO<sub>2</sub> retrievals. *Atmospheric Measurement Techniques* 5.8: 2081-2094.
- Galatry, L. (1961), Simultaneous effect of Doppler and foreign gas broadening on spectral lines, *Phys. Rev.*, 122, 1218-1223.
- Hansen, J.E., and L.D. Travis (1974), Light Scattering in Planetary Atmospheres, *Space Science Reviews*, 16, 527-610.
- Hartmann, J.-M., H. Tran, and G. Toon (2009), Influence of line mixing on the retrievals of atmospheric CO<sub>2</sub> from spectra in the 1.6 and 2.1  $\mu\text{m}$  regions, *Atmos. Chem. Phys. Disc.* 9,, 4873-4898.
- Hovenier, J. W., and J. F. de Haan (1985), Polarized light in planetary atmospheres for perpendicular directions, *Astron. Astrophys.*, 146, 185–191.
- Kawabata, K., and S. Ueno (1988), The first three orders of scattering in vertically inhomogeneous scattering–absorbing media, *Astrophys. Space Sci.*, 150(2), 327–344.
- Lamouroux, J., H. Tran, A. L. Laraia, R. R. Gamache, L. S. Rothman, I. E. Gordon, J.-M. Hartman (2010), Updated database plus software for line-mixing in CO<sub>2</sub> infrared spectra and their test using laboratory spectra in the 1.5–2.3  $\mu\text{m}$  region, *J. Quantitative Spectroscopy & Radiative Transfer*, 111, 2321-2331.
- Laughner, J.L., Kiel, M., Toon, G., Andrews, A., Roche, S., Wunch, D., and Wennberg, P.O. with the ACE-FTS team, ATom team, and Aircore team, Revised formulation of the TCCON priors in GGG2020, *Atmos. Meas. Tech.*, in prep.
- Le Quéré, C., Raupach, M. R., Canadell, J. G., Marland, G., Bopp, L., Ciais, P., Conway, T. J., Doney, S. C., Feely, R. A., Foster, P., Friedlingstein, P., Gurney, K., Houghton, R. A., House, J. I., Huntingford, C., Levy, P. E., Lomas, M. R., Majkut, J., Metzl, N., Ometto, J. P., Peters, G.P., Prentice, I. C., Randerson, J. T., Running, S. W., Sarmiento, J. L., Schuster, U., Sitch, S., Takahashi, T., Viovy, N., van der Werf, G. R., and Woodward, F. I. (2009), Trends in the sources and sinks of carbon dioxide, *Nature Geosci.*, 2, 831–836, doi:10.1038/ngeo689
- Le Quéré, C., C., Andres, R. J., Boden, T., Conway, T., Houghton, R. A., House, J. I., Marland, G., Peters, G. P., van der Werf, G. R., Ahlström, A., Andrew, R. M., Bopp, L., Canadell, J. G., Ciais, P., Doney, S. C., Enright, C., Friedlingstein, P., Huntingford, C., Jain, A. K., Jourdain, C., Kato, E., Keeling, R. F., Klein Goldewijk, K., Levis, S., Levy, P., Lomas, M., Poulter, B., Raupach, M. R., Schwinger, J., Sitch, S., Stocker, B. D., Viovy, N., Zaehle, S., and Zeng, N. (2013), The global carbon budget 1959–2011, *Earth Syst. Sci. Data*, 5, 165–185, doi:10.5194/essd-10 5-165-2013.
- Long, D. A., D. K. Havey, M. Okumura, C. E. Miller, J. T. Hodges (2010), O<sub>2</sub> A-band line parameters to support atmospheric remote sensing, *Journal of Quantitative Spectroscopy and Radiative Transfer*, 111 (14), 2021-2036.
- Long, D. A., D. K. Havey, S. S. Yu, M. Okumura, C. E. Miller, J. T. Hodges (2011), O<sub>2</sub> A-band line parameters to support atmospheric remote sensing. Part II: The rare isotopologues, *Journal of Quantitative Spectroscopy and Radiative Transfer*, 112 (16), 2527-2541.

- Long, D. A., and J. T. Hodges (2012), On spectroscopic models of the O<sub>2</sub> A-band and their impact upon atmospheric retrievals, *Journal of Geophysical Research: Atmospheres*, 117 (D12).
- Magney, T. S., Bowling, D. R., Logand, B. A., Grossmann, K., Stutz, J., Blanken, P. D., Burns, S. P., Cheng, R., Garcia, M. A., Köhler, P., Lopez, S., Parazoo, N. C., Raczka, B., Schimel, D., and Frankenberg, C., (2019), Mechanistic evidence for tracking the seasonality of photosynthesis with solar-induced fluorescence, *Proc. Natl. Acad. Sci. USA*, 11640–11645, doi:10.1073/pnas.1900278116.
- McKenna, D. S., Konopka, P., Grooß, J.-U., Günther, G., Müller, R., Spang, R., Offerman, D. and Orsolini, Y., A new Chemical Lagrangian Model of the Stratosphere (ClaMS), 1, Formulation of advection and mixing, *J. Geophys. Res.*, 107( D16), doi:[10.1029/2000JD000114](https://doi.org/10.1029/2000JD000114), 2002.
- Messerschmidt, J. et al. (2011), Calibration of TCCON column-averaged CO<sub>2</sub>: the first aircraft campaign over European TCCON sites, *Atmos. Chem. Phys.*, 11(21), 10765–10777, doi:10.5194/acp-11-10765-2011.
- Miller, C.E., and L. R. Brown (2004a), Near Infrared Spectroscopy of Carbon Dioxide: <sup>16</sup>O<sup>12</sup>C<sup>16</sup>O Line Positions, *J. Mol. Spectrosc.* 228, 329-354.
- Miller, C. E., M. A. Montgomery, R. M. Onorato, C. Johnstone, T. P. McNicholas, B. Kovaric and L. R. Brown (2004b), Near Infrared Spectroscopy of Carbon Dioxide. II. <sup>16</sup>O<sup>13</sup>C<sup>16</sup>O and <sup>16</sup>O<sup>13</sup>C<sup>18</sup>O Line Positions, *J. Mol. Spectrosc.* 228, 355-374.
- Miller, C. E., L. R. Brown, R. A. Toth D. C. Benner, and V. M. Devi (2005), Spectroscopic challenges for high accuracy retrievals of atmospheric CO<sub>2</sub> and the Orbiting Carbon Observatory (OCO) experiment, *Comptes Rendus Physique* 6, 876-887.
- Miller, C. E., D. Crisp, P. L. DeCola, S. C. Olsen, J. T. Randerson, A. M. Michalak, A. Alkhaled, P. Rayner, D. J. Jacob, P. Suntharalingam, D. B. A. Jones, A. S. Denning, M. E. Nicholls, S. C. Doney, S. Pawson, H. Boesch, B. J. Connor, I. Y. Fung, D. O'Brien, R. J. Salawitch, S. P. Sander, B. Sen, P. Tans, G. C. Toon, P. O. Wennberg, S. C. Wofsy, Y. L. Yung, and R. M. Law (2007), “Precision requirements for space-based XCO<sub>2</sub> data”, *J. Geophys. Res.* 112, D10314, [doi:10.1029/2006JD007659].
- Mlawer, E. J., V. H. Payne, J.-L. Moncet, J. S. Delamere, M. A. Alvarado, D. C. Tobin (2012), Development and recent evaluation of the MT\_CKD model of continuum absorption, *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 370 (2012), 2520-2556.
- Mlawer, E. J., D., G., and Payne, V. H. (2014), Presentation given at the 2014 OCO-2 Science Team Meeting. See: [http://rtweb.aer.com/continuum\\_frame.html](http://rtweb.aer.com/continuum_frame.html)
- Natraj, V., and R. J. D. Spurr (2007), A fast linearized pseudo-spherical two orders of scattering model to account for polarization in vertically inhomogeneous scattering-absorbing media, *J. Quant. Spectrosc. Rad. Tran.*, 107, 263–293.



- Natraj, V., H. Boesch, R. J. D. Spurr, and Y. L. Yung (2008), Retrieval of X<sub>CO2</sub> from simulated Orbiting Carbon Observatory measurements using the fast linearized R-2OS radiative transfer model, *J. Geophys. Res.*, 113, D11212, doi: 10.1029/2007JD009017.
- Niro, F., T. von Clarmann, K. Jucks, and J.-M. Hartmann (2005a), Spectra calculations in central and wing regions of CO<sub>2</sub> IR bands between 10 and 20 m. III: atmospheric emission Spectra, *J. Quant. Spectrosc. Radiat. Transfer* 90, 61–76.
- Niro, F., K. Jucks, and J.-M. Hartmann (2005b), Spectra calculations in central and wing regions of CO<sub>2</sub> IR bands. IV: software and database for the computation of atmospheric spectra, *J. Quant. Spectrosc. & Radiat. Transfer* 95, 469–481.
- O’Brien, Denis M., Igor Polonsky, Chris O’Dell, Akihiko Kuze, Nobuhiro Kikuchi, Yukio Yoshida, and Vijay Natraj (2013), Testing the Polarization Model for TANSO-FTS on GOSAT Against Clear-Sky Observations of Sun Glint Over the Ocean. *IEEE Trans. on Geoscience and Remote Sensing*, 51 (12), 5199–5209.
- O’Dell, C.W. (2010), Acceleration of multiple-scattering, hyperspectral radiative transfer calculations via low-streams interpolation, *J. Geophys. Res.*, 115, D10206.
- O’Dell, C. W., B. Connor, H. Bösch, D. O’Brien, C. Frankenberg, R. Castano, M. Christi et al. (2012). The ACOS CO<sub>2</sub> retrieval algorithm–Part 1: Description and validation against synthetic observations. *Atmospheric Measurement Techniques* 5 (1), 99-121.
- O’Dell, C. W., Eldering, A., Wennberg, P. O., Crisp, D., Gunson, M. R., Fisher, B., Frankenberg, C., Kiel, M., Lindqvist, H., Mandrake, L., Merrelli, A., Natraj, V., Nelson, R. R., Osterman, G. B., Payne, V. H., Taylor, T. E., Wunch, D., Drouin, B. J., Oyafuso, F., Chang, A., McDuffie, J., Smyth, M., Baker, D. F., Basu, S., Chevallier, F., Crowell, S. M. R., Feng, L., Palmer, P. I., Dubey, M., García, O. E., Griffith, D. W. T., Hase, F., Iraci, L. T., Kivi, R., Morino, I., Notholt, J., Ohyama, H., Petri, C., Roehl, C. M., Sha, M. K., Strong, K., Sussmann, R., Te, Y., Uchino, O., and Velazco, V. A., (2018), Improved retrievals of carbon dioxide from Orbiting Carbon Observatory-2 with the version 8 ACOS algorithm, *Atmos. Meas. Tech.*, 11, 6539-6576, <https://doi.org/10.5194/amt-11-6539-2018>, 2018.
- Oyafuso, F., Payne, V. H., Drouin, B. J., Devi, V. M., Benner, D. C., Sung, K., Yu, S., Gordon, I. E., Kochanov, R., Tan, Y., Crisp, D., Mlawer, E. J., and Guillaume, A. (2017), High accuracy absorption coefficients for the Orbiting Carbon Observatory-2 (OCO-2) mission: Validation of updated carbon dioxide cross-sections using atmospheric spectra, *Journal of Quantitative Spectroscopy & Radiative Transfer*, 203, 213–223, doi:10.1016/j.jqsrt.2017.06.012.
- Palmer, P. I., Feng, L., Baker, D., Chevallier, F., Bösch, H., and Somkuti, P., (2019), Net carbon emissions from African biosphere dominate pan-tropical atmospheric CO<sub>2</sub> signal, *Nature Communications*, 10, 3344, doi:10.1038/s41467-019-11097-w.
- Predoi-Cross, A., K. Hambrook, R. Keller, C. Povey, I. Schofield, D. Hurtmans, H. Over, and G. Ch. Mellau (2008a), Spectroscopic lineshape study of the self-perturbed oxygen A-band, *J. Mol. Spectrosc.* 248, 85-110.

- Predoi-Cross, A., C. Holladay, H. Heung, J.-P. Bouanich, G. Ch. Mellau, R. Keller, and D. R. Hurtmans (2008b), Nitrogen-broadened lineshapes in the oxygen A-band: Experimental results and theoretical calculations, *J. Mol. Spectrosc.* 251, 159–175.
- Predoi-Cross, A., A.R.W. McKellar, D. C. Benner, V. M. Devi, R. R. Gamache, C. E. Miller, R. A. Toth and L. R. Brown (2009), Temperature dependences for air-broadened Lorentz half width and pressure-shift coefficients in the 30013←00001 and 30012←00001 bands of CO<sub>2</sub> near 1600 nm, *Can. J. Phys.* (in press).
- Rahman, H., Pinty, B. & Verstraete, M. M. Coupled surface-atmosphere reflectance (CSAR) model: 2. Semiempirical surface model usable with NOAA advanced very high resolution radiometer data. *Journal of Geophysical Research: Atmospheres* 98, 20791–20801 (1993).
- Rayner, P. J. and D. M. O'Brien (2001), “The utility of remotely sensed CO<sub>2</sub> concentration data in surface source inversions,” *Geophys. Res. Lett.*, 28, 175–178.
- Reichler, T., Dameris, M., Sausen, R. Determining the tropopause height from gridded data, *Geo. Res. Let.* 30, 2042, doi:10.1029/2003GL018240, 2003.
- Robichaud, D. J., J. T. Hodges, L. R. Brown, D. Lisak, M. Okumura, and C. E. Miller (2008a), Experimental Intensity and Line Shape Parameters of the Oxygen A-Band Using Frequency-Stabilized Cavity Ring-Down Spectroscopy, *J. Mol. Spectrosc.* 248, 1–13.
- Robichaud, D. J., J. T. Hodges, P. Masłowski, L. Y. Yeung, M. Okumura, C. E. Miller, and L. R. Brown (2008b), High-Accuracy Transition Frequencies for the O<sub>2</sub> A-Band, *J. Mol. Spectrosc.* 251, 27–43.
- Robichaud, D. J., L. Y. Yeung, D. A. Long, D. K. Havey, J. T. Hodges, M. Okumura, C. E. Miller, and L. R. Brown (2009), Experimental Line Parameters of the  $b^1\Sigma_g^+ \leftarrow X^3\Sigma_g^-$  Band of Oxygen Isotopologues at 760 nm Using Frequency-Stabilized Cavity Ring-Down Spectroscopy, *J. Phys. Chem. A*, 2009, 113 (47), pp 13089–13099, doi: 10.1021/jp901127h.
- Rodgers, C. (2000) *Inverse Methods for Atmospheric Sounding: Theory and Practice*. World Scientific Publishing Co Pte Ltd.
- Rodgers, C. D., and B. J. Connor (2003), Intercomparison of remote sounding instruments, *J. Geophys. Res.*, 108(D3), 4116, doi: 10.1029/2002JD002299.
- Rothman L. S., I. E. Gordon, A. Barbe, D. C. Benner, P. F. Bernath, M. Birk, V. Boudon, L. R. Brown, A. Campargue, J.-P. Champion, K. Chance, L. H. Coudert, V. Dana, V. M. Devi, S. Fally, J.-M. Flaud, R. R. Gamache, A. Goldman, D. Jacquemart, I. Kleiner, N. Lacome, W. J. Lafferty, J.-Y. Mandin, S. T. Massie, S. Mikhailenko, C. E. Miller, N. Moazzen-Ahmadi, O. V. Naumenko, A. Nikitin, J. Orphal, A. Predoi-Cross, V. Perevalov, A. Perrin, C. P. Rinsland, M. Rotger, M. Šimečková, M. A. H. Smith, K. Sung, S. Tashkun, J. Tennyson, R. A. Toth, A. C. Vandaele, and J. Vander Auwera (2009), The HITRAN 2008 molecular spectroscopic database, *J. Quant. Spectrosc. Radiat. Transfer* 110, 533–572.
- Rothman, L.S., I.E. Gordon, Y. Babikov, A. Barbe, D. Chris Benner, P. F. Bernath, M. Birk, L. Bizzocchi, V. Boudon, L. R. Brown, A. Campargue, K. Chance, E. A. Cohen, L. H. Coudert, V. M. Devi, B. J. Drouin, A. Fayt, J.-M. Flaud, R. R. Gamache, J. J. Harrison, J.-M. Hartmann, C. Hill, J. T. Hodges, D. Jacquemart, A. Jolly, J. Lamouroux, R. J. LeRoy, G. Li, D. A. Long p,

- O.M.Lyulin, C.J.Mackie, S.T.Massie, S.Mikhailenko, H.S.P.Müller, O.V.Naumenko, A.V.Nikitin, J.Orphal, V.Perevalov, A.Perrin, E.R.Polovtseva, C.Richard, M.A.H.Smith V, E.Starikova, K.Sung, S. Tashkun, J.Tennyson, G.C.Toon, V.I.G.Tyuterev, G.Wagner (2013), The HITRAN 2012 molecular spectroscopic database, *J. Quant. Spectrosc. Radiat. Transfer*, 130, 4.
- Salstein, D. A., Ponte, R. M., and Cady-Pereira, K. (2008), Uncertainties in atmospheric surface pressure fields from global analyses, *J. Geophys. Res.*, 113, D14107.
- Schaepman-Strub, G., Schaepman, M. E., Painter, T. H., Dangel, S. & Martonchik, J. V. Reflectance quantities in optical remote sensing—definitions and case studies. *Remote Sensing of Environment* 103, 27–42 (2006).
- Schutgens, A. J., Tilstra, L. G., Stammes, P. and Breon, F.-M. (2004), On the relationship between Stokes parameters  $Q$  and  $U$  of atmospheric ultraviolet/visible/near-infrared radiation, *J. Geophys. Res.*, 109 (D09205), doi:10.1029/2003JD004081.
- Slijkhuis, S. (1998), SCIAMACHY level 0 to 1c processing algorithm theoretical basis document, Tech. Note ENV-ATB-DLR-SCIA-0041, Dtsch. Zent. für Luft- und Raumfahrt, Oberpfaffenhofen, Germany.
- Slijkhuis, S. (2000), Calculation of polarisation from Rayleigh single scattering, Rep. ENV-TN-DLR-SCIA-0043, Dtsch. Zent. für Luft- und Raumfahrt, Oberpfaffenhofen, Germany.
- Spurr, R. J. D., T.P. Kurosu, and K.V. Chance (2001), A linearized discrete ordinate radiative transfer model for atmospheric remote-sensing retrieval, *J. Quant. Spectrosc. Rad. Tran.*, 68(6), 689–735.
- Spurr, R. J. D. (2002), Simultaneous derivation of intensities and weighting functions in a general pseudo-spherical discrete ordinate radiative transfer treatment, *J. Quant. Spectrosc. Rad. Tran.*, 75(2), 129–175.
- Spurr, R. J. D. (2006), VLIDORT: a linearized pseudo-spherical vector discrete ordinate radiative transfer code for forward model and retrieval studies in multilayer multiple scattering media, *J. Quant. Spectrosc. Rad. Tran.*, 102(2), 316–342.
- Sung, K., L. R. Brown, R. A. Toth, and T. J. Crawford (2009), FT-IR measurements of H<sub>2</sub>O-broadened half-widths of CO<sub>2</sub> at 4.3  $\mu$ m, *Can. J. Phys.* (in press).
- Taylor, T.E., C.W. O'Dell, D.M. O'Brien, N. Kikuchi, T. Yokota, T.Y. Nakajima, H. Ishida, D. Crisp, and T. Nakajima, Comparison of Cloud-Screening Methods Applied to GOSAT Near-Infrared Spectra, *IEEE Trans. Geosci. Remote Sens.*, 50, 295–309, 2012. doi: 10.1109/TGRS.2011.2160270.
- Taylor, T. E., O'Dell, C. W., Frankenberg, C., Partain, P., Cronk, H. Q., Savtchenko, A., Nelson, R. R., Rosenthal, E. J., Chang, A. Y., Fisher, B., Osterman, G., Pollock, R. H., **Crisp, D.**, Eldering, A., and Gunson, M. R: Orbiting Carbon Observatory-2 (OCO-2) cloud screening algorithms; validation against collocated MODIS and CALIOP data, *Atmos. Meas. Tech.*, 9, 973–989, 2016, <http://www.atmos-meas-tech.net/9/973/2016/> doi:10.5194/amt-9-973-2016.

- Taylor, T. E., Eldering, A., Merrelli, A., Kiel, M., Somkuti, P., Cheng, C., Rosenberg, R., Fisher, B., Crisp, D., Basilio, R., Bennett, M., Cervantes, D., Chang, A., Dang, L., Frankenberg, C., Haemerle, V. R., Keller, G. R., Kurosu, T., Laughner, J. L., Lee, R., Marchetti, Y., Nelson, R., O'Dell, C., Osterman, G., Pavlick, R., Roehl, C., Schneider, R., Spiers, G., To, C., Wells, C., Wennberg, P. O., Yelamanchili, A., Yu, S.: OCO-3 early mission operations and initial (vEarly) Xco<sub>2</sub> and SIF retrievals., *Remote Sensing of Environment*, 251, 112032, 2020. doi:10.1016/j.rse.2020.112032.
- Thompson, D. R., D. C. Benner, L. R. Brown, D. Crisp, V. M. Devi, Y. Jiang, V. Natraj, F. Oyafuso, K. Sung, D. Wunch, R. Castano, C. E. Miller (2012), Atmospheric validation of high accuracy CO<sub>2</sub> absorption coefficients for the OCO-2 mission, *Journal of Quantitative Spectroscopy and Radiative Transfer*, 113 (17), 2265-2276.
- Thuillier, G., M. Hersé, P. C. Simon, D. Labs, H. Mandel, D. Gillotay, and T. Foujols (2003), The solar spectral irradiance from 200 to 2400 nm as measured by the SOLSPEC spectrometer from the ATLAS 1-2-3 and EURECA missions, *Solar Physics*, 214(1): 1-22.
- Toon, G.C., J.-F. Blavier, B. Sen, R.J. Salawitch, G.B. Osterman, J. Notholt, M. Rex, C.T. McElroy, and J.M. Russell III (1999), Ground-based observations of Arctic O<sub>3</sub> loss during spring and summer 1997, *J. Geophys. Res.*, 104(D21), 26497-26510.
- Toth, R. A., L. R. Brown, C. E. Miller, V. M. Devi, and D. C. Benner (2006a), Line strengths of <sup>12</sup>C<sup>16</sup>O<sub>2</sub>: 4550-7000 cm<sup>-1</sup>, *J. Mol. Spectrosc.* 239, 221-242.
- Toth, R. A., L.R. Brown, C.E. Miller, V. M. Devi, and D. C. Benner (2006b), Self-broadened widths and shifts of CO<sub>2</sub>, *J. Mol. Spectrosc.* 239, 243-271.
- Toth, R. A., C. E. Miller, L. R. Brown, V. M. Devi, and D. C. Benner (2007a), Line positions and strengths of <sup>16</sup>O<sup>12</sup>C<sup>18</sup>O, <sup>18</sup>O<sup>12</sup>C<sup>18</sup>O and <sup>17</sup>O<sup>12</sup>C<sup>18</sup>O between 2200 and 7000 cm<sup>-1</sup>, *J. Mol. Spectrosc.* 243, 43-61.
- Toth, R. A., C. E. Miller, V. M. Devi, D. C. Benner, and L. R. Brown (2007b), Air-broadened width and pressure shift coefficients of <sup>12</sup>C<sup>16</sup>O<sub>2</sub>: 4700 – 7000 cm<sup>-1</sup>, *J. Mol. Spectrosc.* 246, 133-157.
- Toth, R. A., C. E. Miller, L. R. Brown, V. M. Devi, and D. C. Benner (2008a), Line strengths of <sup>16</sup>O<sup>13</sup>C<sup>16</sup>O, <sup>16</sup>O<sup>13</sup>C<sup>18</sup>O, <sup>16</sup>O<sup>13</sup>C<sup>17</sup>O and <sup>18</sup>O<sup>13</sup>C<sup>18</sup>O between 2200 and 6800 cm<sup>-1</sup>, *J. Mol. Spectrosc.* 251, 64-89.
- Toth, R. A., L. R. Brown, C. E. Miller, V. M. Devi, and D. C. Benner (2008b), Spectroscopic database of CO<sub>2</sub> line parameters: 4300 – 7000 cm<sup>-1</sup>, *J. Quant. Spectrosc. Radiat. Transfer* 109, 906-921.
- Tran, H., C. Boulet, and J.-M. Hartmann (2006), Line mixing and collision-induced absorption by oxygen in the A band: Laboratory measurements, model, and tools for atmospheric spectra computations, *J. Geophys. Res.* 111, D15210, doi: 10.1029/2005JD006869.
- Tran, H., and J.-M. Hartmann (2008), An improved O<sub>2</sub> A band absorption model and its consequences for retrievals of photon paths and surface pressures, *J. Geophys. Res.* 113, D18104, doi: 10.1029/2008JD010011.
- van de Hulst, H. C. (1981), *Light Scattering by Small Particles*, Dover, Mineola, N. Y.

- Washenfelder, R. A., G. C. Toon, J.-F. Blavier, Z. Yang, N. T. Allen, P. O. Wennberg, S. A. Vay, D. M. Matross and B. C. Daube (2006), Carbon dioxide column abundances at the Wisconsin Tall Tower site, *J. Geophys. Res.* 111, D22305, doi: 10.1029/2006JD007154.
- Wiscombe, W. J. (1977), The delta-M method: rapid yet accurate radiative flux calculations for strongly asymmetric phase functions, *J. Atmos. Sci.*, 34 (9), 1408-1422.
- World Meteorological Organization (WMO), Meteorology A Three-Dimensional Science: Second Session of the Commission for Aerology, *WMO Bulletin IV*(4), WMO, Geneva, 134-138, 1957.
- Wunch, D. , G. C. Toon, P. O. Wennberg, S. C. Wofsy, B. B. Stephens, M. L. Fischer, O. Uchino, J. B. Abshire, P. Bernath, S. C. Biraud, J.-F. L. Blavier, C. Boone, K. P. Bowman, E. V. Browell, T. Campos, B. J. Connor, B. C. Daube, N. M. Deutscher, M. Diao, J. W. Elkins, C. Gerbig, E. Gottlieb, D. W. T. Griffith, D. F. Hurst, R. Jimenez, G. Keppel-Aleks, E. Kort, R. Macatangay, T. Machida, H. Matsueda, F. Moore, I. Morino, S. Park, J. Robinson, C. M. Roehl, Y. Sawa, V. Sherlock, C. Sweeney, T. Tanaka, and M. A. Zondlo, (2010), Calibration of the total carbon column observing network using aircraft profile data, *Atmos. Meas. Tech. Discuss.*, 3, 2603–2632, doi: 10.5194/amtd-3-2603-2010.
- Wunch, D., G. C. Toon, J.F.L. Blavier, R.A. Washenfelder, J. Notholt, B. J. Connor, D.W.T. Griffith, V. Sherlock, P.O. Wennberg (2011) The Total Carbon Column Observing Network, *Phil. Trans. Of the Royal Society A-Mathematical Physical and Engineering Sciences*, 369, 2087-2112, DOI: 10.1098/rsta.2010.0240.
- Yang, Z., P. O. Wennberg, R. P. Cageao, T. J. Pongetti, G. C. Toon, and S. P. Sander (2005), Ground-based photon path measurements from solar absorption spectra of the O<sub>2</sub> A-band, *J. Quant. Spectrosc. Radiat. Transfer* 90, 309–321.
- Young, A.T. (1980), Revised depolarization corrections for atmospheric extinction, *Applied Optics*, 19 (20), 3427-3428.
- C. Zhao, P.P. Tans, and K.W. Thoning (1997), A high precision manometric system for absolute calibrations of CO<sub>2</sub> in dry air, *J. Geophys. Res.* 102, 5885-5894.
- Zong, J. (2008), Application of Global High-Resolution DEM in the OCO Mission, in: International Archives of the Photogrammetry, Remote Sensing, and Spatial Information Sciences, Vol. XXXVII, Part B4, Beijing.

## 6 Acronyms

2D, 4D	Two- and four-dimensional
2OS	First two orders of scattering (model)
A-Train	Earth Observing System Afternoon Constellation
ABSCO	Absorption coefficients
ACS	Attitude control system
AOD	Aerosol optical depth
ATBD	Algorithm theoretical basis document
AU	Astronomical Unit
CIA	Collision-induced absorption
CO <sub>2</sub>	Carbon Dioxide
CRDS	Cavity ring-down spectrometer
EOF	Empirical orthogonal function
FOV	Field of view
FP	Full physics
FPA	Focal plane array
FTS	Fourier transform spectrometer
GEOS FP-IT	Global Earth Observing System Forward Processing for Instrument Teams
GOSAT	Greenhouse Gases Observing Satellite
GSFC	NASA Goddard Space Flight Center
H <sub>2</sub> O	Water
HITRAN	High Resolution Transmission database
ILS	Instrument line shape
IMAP-DOAS	Iterative maximum <i>a posteriori</i> differential optical absorption spectroscopy
JPL	NASA Jet Propulsion Laboratory
L (0,1..)	Level 0, Level 1, etc. (data product)
LIDORT	Linearized discrete ordinate radiative transfer
LSI	Low-streams interpolation
MERRA	Modern-Era Retrospective Analysis for Research and Applications
MODIS	Moderate-Resolution Imaging Spectroradiometer
NASA	National Aeronautics and Space Administration
NIR	Near-infrared
NIST	National Institute of Standards and Technology (Maryland)
O <sub>2</sub>	Oxygen
OCO	Orbiting Carbon Observatory
PGE	Product Generation Executable
ppm	parts per million
RH	Relative humidity
RT	Radiative transfer (model)
RTE	Radiative transfer equation
SDOS	Science Data Operations System

SIF	Solar-induced fluorescence
SOLSPEC	Solar Spectrum (instrument)
SNR	Signal-to-noise ratio
SWIR	Short-wave-infrared
TCCON	Total Carbon Column Observing Network
TOA	Top of atmosphere
VLIDORT	Vector LIDORT
WMO	World Meteorological Organization
X <sub>CO2</sub>	Column-averaged CO <sub>2</sub> dry air mole fraction